



How Object-Based Storage Can Meet 5 Big Data Challenges

Introduction



The rise of big data shows no signs of slowing down. Companies are storing massive amounts of data that are reaching previously unheard of levels. According to the IDC, 80% of new cloud applications will be “big data intensive,” potentially requiring petabytes of storage.¹ New technology trends are contributing to data growth. Information from mobile devices, sensors on objects in the Internet of Things, and social media alone funnels tons of data into your business.

The volume, variety, and velocity of big data strains legacy storage infrastructures. In traditional infrastructures, data coming from separate sources gets siloed. As a result, it is only available to different departments. With data silos, it’s hard to see the big picture.

Given its hierarchical organizational system, traditional file storage can make accessing data inefficient, like searching for a book in a library. You need the equivalent of a Dewey Decimal number to find the information you need. This method of accessing data makes working with unstructured data particularly difficult.

The huge volumes of data streaming into companies are the key to making strategic business decisions. To leverage the profitable insights hidden in data, you need storage that will make data easier to access and interpret.

This eBook explores 5 challenges created by big data. Understanding these challenges will help you decide if object-based storage is the best solution for meet your big data demands.

¹ Fortune | <http://fortune.com/2015/10/05/ibm-buys-cleversafe/>

Scaling for Data Growth

Business Insider reports that data is doubling in size every 2 years.² Meeting the storage needs of big data can be expensive due to required energy, cooling, and floor space. When faced with the challenge of storing petabytes of data, the typical reaction is to provision more hardware to meet demand. The result: server sprawl.

The more servers contained in the data center, the more energy required to prevent downtime caused by overheating. Hard disk drives require energy for not only cooling, but also for keeping the disks spinning.

Scaling on-premises storage is time-consuming, requiring the purchase and installation of hardware. When workloads change, you can't scale-up on demand—risking the development of bottlenecks. Unpredictable workload demands may also derail your budget plans.

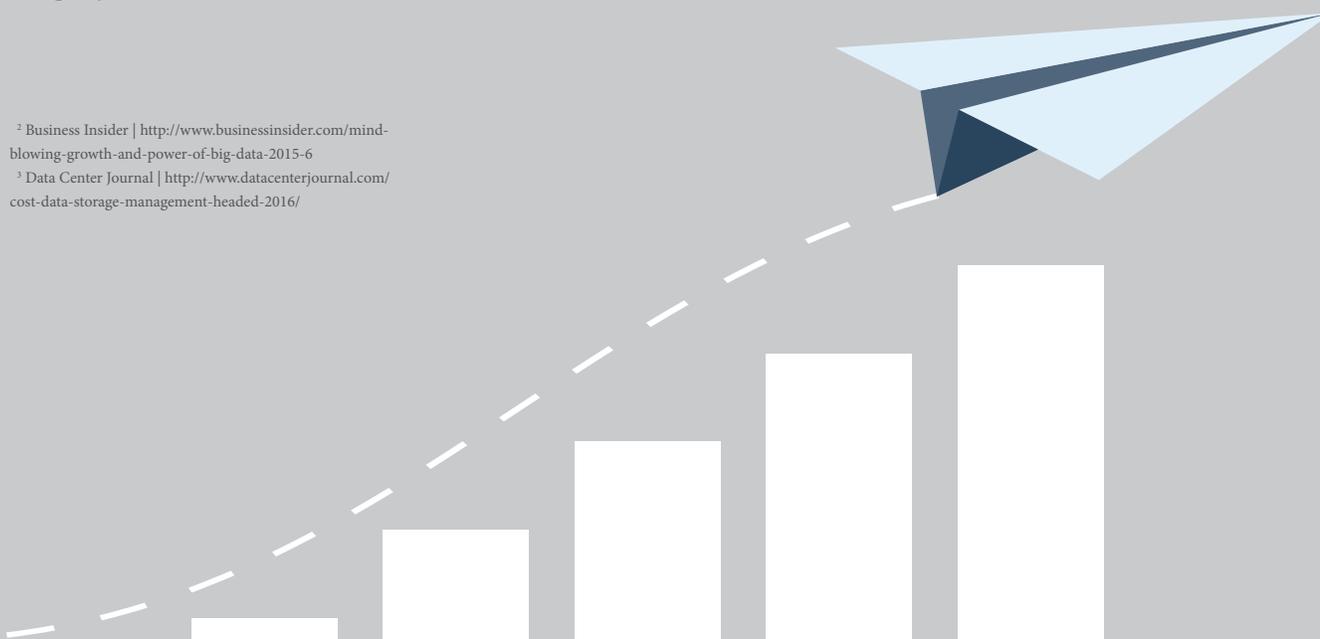
The Capacity Guessing Game

Capacity needs are difficult to estimate. Many IT leaders over-provision storage because they are afraid they will fall short of capacity demands. If this extra storage isn't needed, it can be a waste of money. On the other hand, if you under-provision, your users complain of network and application performance problems. Ultimately, you may need to suddenly shut down your systems to add more storage.

Meeting capacity requires more economic and efficient means of choosing the right method of storage and using the right data management strategies, such as compression and automatic storage tiering. Data that is used infrequently needs to be moved to lower-performing storage. According to Data Center Journal, 60% of companies' data can be stored on slower, high-capacity storage.³

² Business Insider | <http://www.businessinsider.com/mind-blowing-growth-and-power-of-big-data-2015-6>

³ Data Center Journal | <http://www.datacenterjournal.com/cost-data-storage-management-headed-2016/>



Making Sense of Unstructured Data

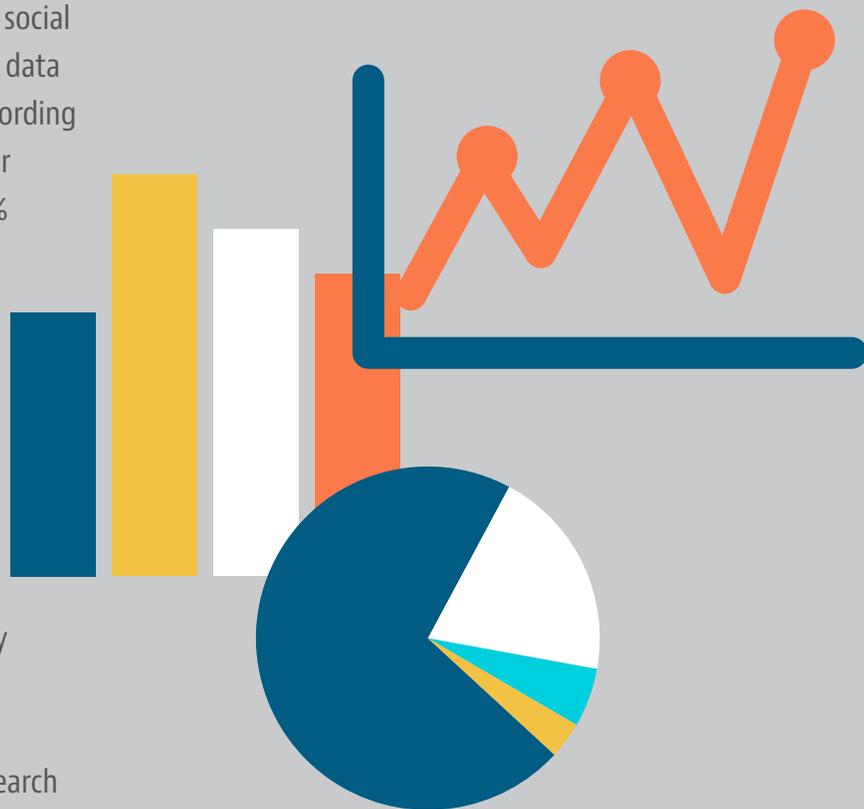
The massive volume of information is not the only challenge big data presents. Big data comes in a variety of forms, including text, video, audio, and image. This unstructured data comes from sources like social networks and mobile devices. As the overall volume of data grows, the amount of unstructured data increases. According to IDG, unstructured data increases at a rate of 62% per year.⁴ In addition, Forbes reports that a staggering 80% of data is unstructured.⁵ Your business can't afford to ignore it.

Getting Lost in Unstructured Data

Unstructured data cannot be easily organized and can be difficult to search since it cannot be stored like structured data in relational databases with rows and columns. The information can only be identified by filename and extension.

Companies need to find better ways to organize and search for unstructured data because it reveals more detailed, qualitative information about the significance of structured data. For instance, unstructured data reveals the motivations behind decisions that customers and clients make rather than providing figures.

Understanding these end-user motivations and sentiments can help your business meet client needs more effectively and design a more customizable customer experience for greater loyalty. Customers and users comment on problems they are having with service or quality, spotlighting changes you can make to business processes.



⁴ Beta News | <http://betanews.com/2016/04/13/crowdsourcing-platform-creates-insights-from-unstructured-data/>

⁵ Forbes | <http://www.forbes.com/sites/steveandriole/2015/03/05/the-other-side-of-analytics/2/#284067d74ebe>

Maintaining Compliance

Database compliance is a particular challenge for organizations in the financial and healthcare industries. To meet compliance regulations, companies need to maintain active archives whether data is being used or not. Companies also need to maintain auditable records of sensitive data access and transmission.

Financial Compliance

In finance, the Dodd-Frank Act and Sarbanes-Oxley Act require that companies maintain records of transactions for 5 to 7 years. These records include auditing and real-time transaction monitoring reports that need to be available on demand.

The Gramm-Leach-Bliley Act requires that financial organizations explain their information-sharing processes with their clients in addition to protecting their personal information. Financial companies need to protect account information, cardholder data, and transaction records while in transit and at rest.

Maintaining compliance makes financial sense. The Ponemon Institute estimates the cost of non-compliance is 2.65x higher than maintaining compliance.⁶

Healthcare Compliance

In healthcare, the Health Insurance Portability and Accountability Act (HIPAA) regulates the protection of private and personally identifiable information in electronic health records (EHR).⁷ Failure to comply can result in fines as high as \$50,000 per offense, even if no wrongdoing was intended.⁸ For HIPAA compliance, data needs to be accessible, yet secure. Patients need to be able to access their data on request, but it must be protected from unauthorized access. If a breach occurs, affected patients need to be notified in a timely manner.

Despite all of these strict security regulations, the Office of Civil Rights reported that there were 253 data breaches at healthcare institutions in 2015, which resulted in the loss of 112 million records.⁹

In particular, EHR can be difficult to search and protect. Like much of big data, EHR consists of unstructured data like doctor's notes. During medical treatment, this data needs to be transferred between members of a medical team, making it vulnerable to being lost or stolen.

⁶ The Ponemon Institute | http://www.ponemon.org/local/upload/file/True_Cost_of_Compliance_Report_copy.pdf

⁷ US Department of Health and Human Services | <http://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html>

⁸ HIPAA | <http://www.hipaa.com/the-reality-of-hipaa-violations-and-enforcement/>

⁹ Forbes | <http://www.forbes.com/sites/danmunro/2015/12/31/data-breaches-in-healthcare-total-over-112-million-records-in-2015/#24e812b07fd5>

Accessing Real-time Insights

Data needs to be accessed quickly for actionable insights and real-time decision making. Real-time analytics helps make use of data within a minute of the data being generated. Accessing and processing unstructured data can be a stumbling block in the real-time analytics process. Real-time analytics lend themselves to the batch-oriented databases used for structured data, but not to unstructured data.

To make the most of real-time analytics, your database needs to store unstructured data so it can be accessed quickly. This is especially important because unstructured data holds the key to gaining a complex understanding of both your business and your customers.

The Benefits of Real-time Analytics

In terms of the customer relationship, real-time access ensures the relevance of unstructured data coming in from customer interactions. Responding quickly to customer needs is crucial in this customer-centric era. Customers expect to be able to get or access what they want when they want it. Real-time analytics are the key to responding to trends in customer or user behavior.

Within your business, real-time analytics enables you to solve problems immediately. Sensors from equipment provide information about which business processes are running smoothly and which are making errors. Identifying and solving a problem immediately can save your company money and help you gain an edge over your competitors.



Harnessing Cognitive Power with the New Wave of Analytics

Forrester found that companies use only 25% of their unstructured data to gain insights.¹⁰ A lack of the appropriate analytic tools may account for this missed opportunity. Data analytics has progressed to include predictive analytics that forecasts trends and prescriptive analytics that proposes courses of action. But neither approach explores the full potential of unstructured data.

Master Unstructured Data with Cognitive Analytics

Understanding the full significance of unstructured data requires cognitive analytics. IDC predicts that by 2020, half of business analytics will build prescriptive analytics on cognitive computing.¹¹ Cognitive analytics uses machine learning and natural language processing to make use of data that was previously unusable. For instance, natural language processing can draw associations between topics in textual documents like those found in unstructured data.

Cognitive analytic systems are able to store and process encyclopedic amounts of information. Using machine learning, the system can learn, reason, and draw hypotheses that can be used to make complex business decisions.

Cognitive analytics is a disruptive trend that shows no sign of stopping. Gartner identified cognitive analytics as a top trend for 2016.¹² In the time that IBM's Watson has been operating, its capabilities have quickly multiplied as it moved from giving the winning responses on *Jeopardy* to being able to use image analytics to see.¹³

¹⁰ Forrester | http://blogs.forrester.com/boris_evelson/15-04-29-expand_your_big_data_capabilities_with_unstructured_text_analytics

¹¹ Forbes | <http://www.forbes.com/sites/gilpress/2015/12/15/6-predictions-for-big-data-analytics-and-cognitive-computing-in-2016/#1f9caa73409e>

¹² Information Week | http://www.informationweek.com/big-data/big-data-analytics/big-data-predictions-for-2016/d/d-id/1323671?image_number=4

¹³ Fast Company | <http://www.fastcompany.com/3049570/tech-forecast/ibms-watson-can-now-see>



Conquer Big Data with Object-based Storage

The challenges presented by big data may seem overwhelming, but there could be a single solution.

[Object-based storage](#) provides the type of organization you need to make use of unstructured data. IDC predicts that the need to store petabytes of data will drive the object storage market to the \$28 billion mark by 2018.¹⁴

[Object-based storage](#) provides a solution to storage bottlenecks by combining scale-out architecture with fast filtering and search. With object-based storage, you can efficiently store and access the most complicated and valuable information that comes into your business: unstructured data.

Data is stored in a flat address space called an object pool, rather than hierarchical systems like batches, files, or columns. Every object has a unique ID that enables applications and users to access data regardless of its location. With object-based storage, you can access information according to what it is, not where it is.

Recently, IBM acquired [Cleversafe](#), transforming it into an IBM Cloud Object Storage Solution. IBM Cloud Object Storage stores data more efficiently for access to real-time data. Your data is kept secure through encryption and a self-healing system that eliminates the need for replication.

As a Premier IBM Business Partner, Vicom gives your business access to the expertise needed to make the most of IBM Cloud Object Storage—formerly [Cleversafe](#).

¹⁴ Fortune | <http://fortune.com/2015/10/05/ibm-buys-cleversafe/>

Is your business ready to win the big data challenge with IBM Cloud Object Storage? **Reach out to Vicom for a free consultation to see how IBM Cloud Object Storage can fit into your business.**

