



White Paper

The ABCs of Big Data – Analytics, Bandwidth and Content

Richard Treadway and Ingo Fuchs, NetApp
March 2012 | WP-7147

EXECUTIVE SUMMARY

Enterprises are entering a new era of scale, where the amount of data processed and stored is breaking down every architectural construct in the storage industry. NetApp delivers solutions that address big data scale through the “Big Data ABCs” – analytics, bandwidth and content – enabling customers to gain insight into massive datasets, move data quickly, and store important content for long periods of time without increasing operational complexity.

TABLE OF CONTENTS

1	ENTERING A NEW ERA OF BIG SCALE	3
2	SOURCES OF BIG SCALE	3
3	SHORTCOMINGS OF CURRENT APPROACHES	3
4	INFRASTRUCTURE BREAKING POINTS	4
5	THE ABCS OF DATA AT SCALE	5
6	SUMMARY	5

LIST OF FIGURES

Figure 1)	Where is your infrastructure breaking?	4
Figure 2)	The NetApp Big Data ABCs: Analytics, Bandwidth, and Content	5

1 ENTERING A NEW ERA OF BIG SCALE

In the 1990s, IT teams were focused on obtaining optimal performance from the key applications and infrastructure of their enterprises. These siloed “systems of record” typically did a good job of keeping track of vital information, but they were very expensive and did not offer sufficient drill-down insight into the data to drive business advantage. In the 2000s, the IT focus shifted to efficiency – how to do more with less. Technologies like virtualization, sharing, and consolidation of the enterprise’s existing infrastructure became the key drivers for IT.

We are now entering a new era of big scale, where the amount of data processed and stored by enterprises is breaking down every architectural construct in the storage industry today. As a result, IT teams are trying to convert these existing systems of record, built back in the 1990s and 2000s, into “systems of engagement” – systems that can efficiently deliver the necessary information, to the right people, in real time, to help them perform more sophisticated analyses and make better business decisions.

Evolving from Systems of Record to Systems of Engagement

Data by itself has no value. Value comes from using the data to drive business results, offer services to customers, and increase revenue. The challenge for scalable storage is to enable these business outcomes from dramatically larger datasets.

2 SOURCES OF BIG SCALE

This massive increase in scale is occurring for a number of reasons. Because of cost pressures, many companies are consolidating their data centers – they can no longer afford for every business unit to have its own IT infrastructure distributed around the globe. The move to cloud computing also contributes to increased scale, aggregating the demand of hundreds of thousands of users onto fewer, centralized systems.

Another source of the increase in scale is the massive growth in machine-generated and user-generated data. Digital technologies are moving to denser media, photos have all gone digital, videos are using higher resolution, and advanced analytics require more storage. Furthermore, machine-generated data from sensor networks, buyer behavior tracking, and other sources contribute to much larger datasets that need to be understood and commercialized. In short, the amount of data is increasing and the data objects themselves are getting bigger. All of these forces together put an enormous amount of scale pressure on existing infrastructures, especially the storage platform. This is what NetApp refers to as the Big Data Challenge.

Where is Big Data Coming From?

Although human-generated data, such as Facebook pictures and Tweets, is getting the most attention in the media, the biggest data growth comes from machine-generated datasets, such as consumer behavior tracking and financial market analyses.

3 SHORTCOMINGS OF CURRENT APPROACHES

Today’s enterprises are finding it difficult to manage the exponential growth in big data. Traditional approaches can’t scale to the level needed to be able to ingest all of the data, analyze it at the speed at

which it arrives, and store the relevant datasets efficiently for extended periods of time. The industry as a whole has started to get a handle on how to manage the increased infrastructure complexity in a virtual world, but handling infrastructure in a scalable world presents some very serious challenges.

Time-to-information is critical for enterprises to derive maximum value from their data. If it takes weeks or months to run an analysis, it won't be timely enough to detect a pattern that may affect the business in an instant. Compliance is also a significant challenge for many enterprises. Regulated organizations may have to keep data for very long periods of time – or forever. And they are required to find the data quickly when needed for reporting or during industry audits.

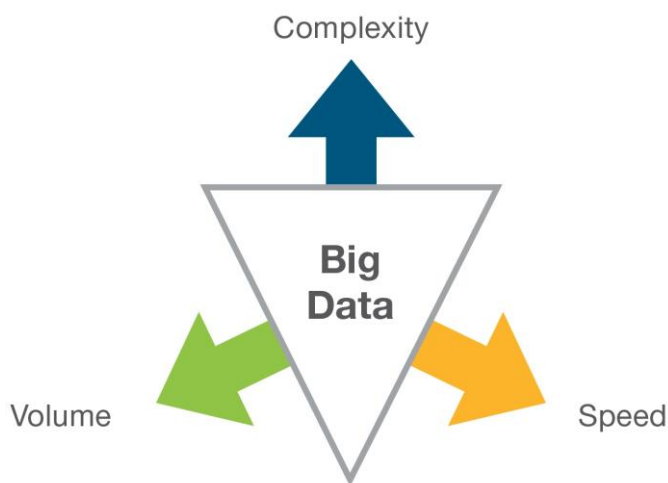
In summary, the Big Data Challenge is all about gaining business advantage – how to obtain the most value for the enterprise from this immense digital universe of information.

4 INFRASTRUCTURE BREAKING POINTS

Big data is breaking today's storage infrastructure along three major axes, as illustrated in Figure 1.

- **Complexity.** Data is no longer just about text and numbers; it's about real-time events and shared infrastructure. The information is now linked, it is high fidelity, and it consists of multiple data types. Applying normal algorithms for search, storage, and categorization is becoming much more complex and inefficient.
- **Speed.** How fast is the data coming in? High-definition video, streaming media over the Internet to player devices, slow-motion video for surveillance – all of these have very high ingestion rates. Businesses have to keep up with the data flow to make the information useful. They also have to keep up with ingestion rates to drive faster business outcomes – or in the military, to save lives.
- **Volume.** All collected data must be stored in a location that is secure and always available. With such high volumes of data, IT teams have to make decisions about what is “too much data.” For example, they might flush all data each week and start all over the following week. But for many applications this is not an option, so more data must be stored longer – without increasing the operational complexity. This can cause the infrastructure to quickly break on the axis of volume.

Figure 1) Where is your infrastructure breaking?

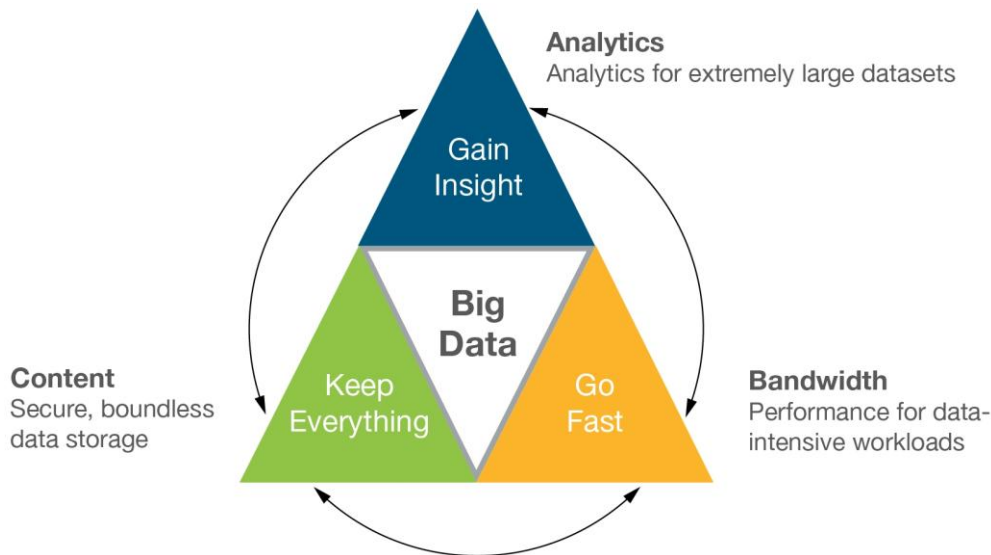


5 THE ABCS OF DATA AT SCALE

NetApp has divided the solution sets for managing data at scale into three main areas, called the “Big Data ABCs” – analytics, bandwidth, and content. As shown in Figure 2, each area has its own specific challenges and unique infrastructure requirements.

- **Analytics.** This solution area focuses on providing efficient analytics for extremely large datasets. Analytics is all about gaining insight, taking advantage of the digital universe, and turning data into high-quality information, providing deeper insights about the business to enable better decisions.
- **Bandwidth.** This solution area focuses on obtaining better performance for very fast workloads. High-bandwidth applications include high-performance computing: the ability to perform complex analyses at extremely high speeds; high-performance video streaming for surveillance and mission planning; and as video editing and play-out in media and entertainment.
- **Content.** This solution area focuses on the need to provide boundless secure scalable data storage. Content solutions must enable storing virtually unlimited amounts of data, so that enterprises can store as much data as they want, find it when they need it, and never lose it.

Figure 2) The NetApp Big Data ABCs: Analytics, Bandwidth, and Content



6 SUMMARY

The new era of scale is breaking existing storage architectures. Enterprises need to ask the following questions: Are there any opportunities for us to take better advantage of our data? What insights can really help our business? Where could we use our data to competitive advantage? What if we could link the trends in buying patterns to people's physical location at a point in time to give them a better

experience? What if we could detect when fraud is about to happen? Can we identify the likely hotspots for failure before it happens?

The list of questions is unlimited. But the answer is always the same. NetApp offers the storage solutions that enable enterprises to take advantage of big data and transform it into greater business value. The universe of data can be an information gold mine. NetApp helps enterprises find the value of this data and turn it into real business advantage.

Your Big Data innovation is based on NetApp

NetApp's Big Data offerings provide a foundation to spark innovation, make better decisions and drive successful business outcomes at the speed of today's business.

NetApp provides no representations or warranties regarding the accuracy, reliability or serviceability of any information or recommendations provided in this publication, or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information in this document is distributed AS IS, and the use of this information or the implementation of any recommendations or techniques herein is a customer's responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. This document and the information contained herein may be used solely in connection with the NetApp products discussed in this document.

Go further, faster®