box UK

White Paper

Seven Steps to Big Data Success

Providing you with:

- An introduction to big data
- Advice on defining a big data strategy
- Best practice recommendations

Introduction

An explosion of connected devices along with systems capable of rapidly processing, storing and analysing large amounts of disparate and often unstructured information has led to 'big data' becoming one of the hottest topics around; no longer limited to the field of science and mathematics but instead a mandatory consideration for leaders across a wide range of sectors, departments and roles.

> **Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization "**
>
> **Wikipedia**

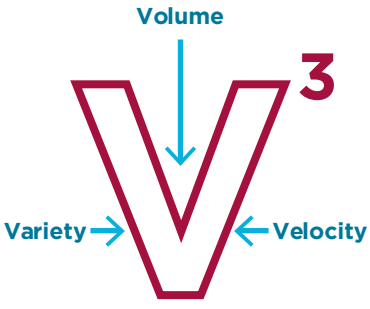
A number of factors have contributed to this trend, not least a decrease in the cost of gathering and storing information that has made the prospect of looking at big data in detail more accessible than ever, and an increase in the volume and variety of devices able to transmit and receive data brought about by the growth in the 'Internet of Things'. This not only enables greater amounts of data to be stored but makes the analysis of this information a vital element in understanding the multi-channel consumer journey and achieving better, faster decision-making.

The potential benefits of big data for organisations are well-recognised; in the UK data-driven firms are, on average, 8% more productive, have 4% higher return on equity, and are 40% more likely to report launching products and services ahead of their non data-driven competitors (source: Nesta). However, with these

opportunities also come challenges; fewer than 10% of organisations believe they are currently in a position to make use of the information they have on customer preferences, behaviour and demands, with 85% of Chief Financial Officers (CFOs) and Chief Information Officers (CIOs) stating that they don't know how to analyse the data they have already collected, 54% citing an inability to identify the data worth collecting as their greatest barrier to success, and 40% saying their greatest challenge is the integration of data analytics into existing systems (source: KPMG). These concerns may help to explain why only 12.5% of organisations have currently implemented big data initiatives (source: IDG).

> **"Big data is a collection of data from traditional and digital sources inside and outside your company that represents a source for on-going discovery and analysis."**
>
> **Forbes**

This white paper looks at some of the key issues organisations must address before putting a big data strategy in place, along with general best practice recommendations to get the greatest business value out of any data gathered.



What is big data?

While the specifics of big data differ depending on who you ask, definitions often feature the so-called 'Three Vs':

- Volume: the amount of data being produced
- Variety: the number of disparate data formats, relationships and sources
- Velocity: the speed with which data can be generated, processed and acted upon

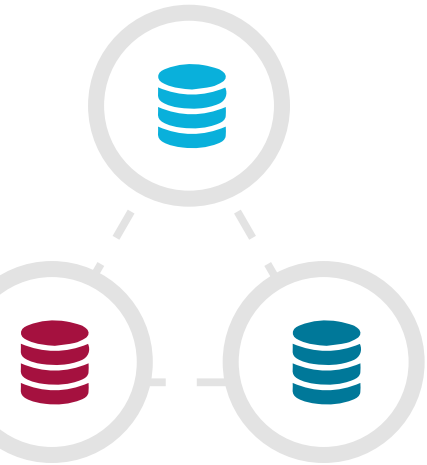
It is generally agreed that possessing high levels of some, or all, of the above qualities makes data ‘big’.

> **“Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization.”**
>
> **Gartner**

In addition, IBM includes a fourth V in its own definition of big data: ‘veracity’, covering the challenges faced when attempting to ensure the accuracy of information received.

1

Join up your view of existing data sources



This section deals with two of big data's four Vs mentioned earlier: volume and variety. The fact that we are producing more data than ever before is undeniable; every day 2.5 billion gigabytes of data is created, enough to fill over 27,000 iPads per minute (source: HM Government), and this is predicted to grow by a factor of 10 by 2020 (source: Wikibon). This has had a direct impact on the amount of information available to organisations; take Procter & Gamble as an example, which now has access to data from over 1500 websites, 1.2 billion emails and 500 different Customer Relationship Management (CRM) programmes (source: BigData-Startups).

These staggering statistics are further complicated by the fact that the vast majority of this 'new-growth' data is unstructured, being sent in formats that cannot be easily and quantifiably stored and analysed such as emails, images, videos and social updates. As well as an increasing variety of data formats, the range of customer touchpoints is also becoming more diverse, with the average person carrying three connected devices on them at all times (source: Sophos). It has also been predicted that by 2020 sensor data will be created by as many as 50 billion connected devices around the world (source: HM Government).

So, how can organisations hope to manage in this brave new world? Thankfully as the volume and variety of data is growing so are the capabilities of solutions providers, with software such as Hadoop, which enables the processing of large, complex data sets, becoming an increasingly popular enterprise solution.

Before selecting the tools you're going to use, however, it is vital to define a data collection and storage

infrastructure. Consider the ways in which this will need to be tailored to your environment and requirements; for example, cloud-based solutions may be well-suited to small organisations but might not provide the speed and security needed for larger enterprises, while attention must also be paid to how you will transition from any legacy systems or processes you have in place.

While the direction of your infrastructure strategy will depend ultimately on your business goals (a consideration covered later in this report), achieving a joined-up view of activity should be a priority for all organisations to help improve the relevance of user interactions and deliver increased engagement and conversion rates. Your infrastructure therefore needs to be able to accommodate the various channels, sources and formats required to build up this picture, as well as have the capability to scale not only as your organisation expands, but across all necessary departments and divisions to eliminate data silos.

2

Ensure you can undertake rapid data analysis

As with volume and variety, the third V, velocity, has been increasing exponentially in recent years with more and more devices, such as monitoring equipment and analytical tools, capable of streaming vast amounts of data continuously. Not only does this present infrastructure challenges surrounding the capture and storage of this data, it also raises important questions regarding how the data can be rapidly queried.

With users increasingly expecting and even demanding content and services be tailored to their specific preferences and buying behaviours, many organisations are starting to look at ways to leverage relevant, up-to-date information in their interactions with their audience; from helping support teams more quickly pinpoint and resolve customer issues to generating relevant cross-selling promotions and offers even as the initial conversation is taking place. However, as we produce a larger volume of data faster than ever before its useful life becomes increasingly limited, giving competitive advantage to those organisations who can most rapidly access and apply the information available to them.

This has created a demand for technology solutions that can process data in fractions of a second, while also ensuring the highest levels of performance, reliability and security. Leading the way in this area are relational NoSQL databases such as MongoDB, which have been designed with speed as a specific concern; allowing fast, intelligent searches to be performed that deliver the understanding needed to improve the user experience exactly when it's required.

3

Validate the accuracy of your data

We come now to the final V, veracity, which is critically important to ensure that you are adhering to any legal requirements surrounding the use of data (covered in more detail later). It also helps you understand what data will best help you generate relevant, actionable insights by reducing unnecessary and unwanted noise and highlighting only meaningful signals of behaviour and intent.

To be confident in the accuracy of your analysis you must be confident in the accuracy of your data, and as such a formal data-cleansing process should be one of the first activities you undertake. Setting up a viable infrastructure as detailed previously will provide you with a collated set of data, enabling de-duplication procedures to be performed to remove any repeat entries liable to mask true patterns and trends.

Make sure also that all your data follows best practice; for example, apply metadata to inform users of the data's provenance, and include guidance as to appropriate usage if necessary. With these valuable details in place, no matter who accesses your data they will be able to immediately understand what it is, where it came from, how old it is and how it can be applied; helping ensure that the correct information is used in each case and improving the quality of any analysis conducted as a result.

Another key factor that may affect confidence in your data is whether it is first-, second- or third-party.

First-party data: data collected and owned by you; for example, from your site, user surveys or CRM data

Second-party data: data about your organisation but

gathered by an external supplier; for example, analytics and social data

Third-party data: data gathered from and provided by an external supplier; for example, purchased segmentation lists

With organisations having the greatest control over their first-party data, it should form a central element of any strategy. After all, information about your own users is likely more valuable than that derived from generic audience segments, and you can be sure about the relevance and recency of the data. Second- and third-party data does however also offer significant value, particularly when identifying opportunities during early-stage decision making. For greatest results, it may be desirable to combine different data sources; for example, exploring the potential effectiveness of various channels with third-party benchmark data then conducting testing and optimisation with your own findings, or assessing internally-collected data to identify high-performing segments then looking to third-party sources to target these further.

4 Deliver real value

So far we have covered the main qualities that can be said to differentiate big data from other kinds of information, and have looked at some of the ways this can be managed through effective collection and processing. Simply 'having' data though isn't a strategy; to make a positive impact on your organisation's performance this data must be strategically applied to speed up OODA (Observe, Orient, Decide, Act) loops and reduce levels of risk throughout the cycle by supporting decision-making with empirical evidence. However, achieving this is a challenge for many organisations, and reports show that while 99% of organisations consider data and analytics to be at least somewhat important to their business strategies, 96% admit that "untapped benefits remain on the table" (source: KPMG); further to this, only 12% of big data projects have board-level sponsors who can help ensure activity remains focused on delivering high-level business benefits (source: MBN).

To leverage big data insights in a strategic manner, they need to be aligned with your business goals. For a truly targeted approach, it's vital that these goals are defined first then used to guide the kinds of data you collect and the ways in which it is employed, rather than assessing the data you already have and trying to craft your strategy around it. This also applies to tools, frameworks and processes; instead of starting with what you have, or even what you think you want, come to your strategy free from preconceptions and existing requirements to help ensure that you focus solely on supporting your core business objectives.

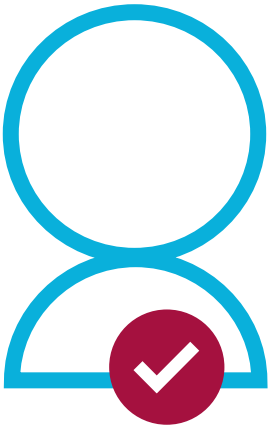
Include all necessary team members in the process, for example individuals who have frequent dealings with your customers such as marketers, support staff and service representatives, as they will be able to provide enhanced insight into what data will be most helpful to positively

impact the service you provide to users.

Remember, too, that although the rise of big data has been facilitated by increasingly sophisticated real-time technologies, traditional skills such as business analysis and strategic planning remain fundamental to developing an understanding of what you want to achieve and creating a robust plan to get you there. As such, ensure your team possesses a balance of technical- and business-focused skills, and also consider how the traditional forms of business intelligence you have access to can complement, support and enhance big data.

5

Put in place the skills to extract information



Addressing the issues raised in the previous sections should help ensure your data is well-organised and valuable, and provide you with a clearly defined understanding of what you want to get from it. How then do you join up these two elements to unlock the power of the data you hold?

Applying visualisation techniques (such as representing data graphically or pictorially) allows you to tell stories and generate meaningful insights that can be easily understood across your organisation and applied to day-to-day business activities and decisions. Achieving this, however, cannot be accomplished with just technology and process, but requires people; specifically, people who are skilled in a wide and diverse range of areas including data management, analytics and business intelligence.

The demand for these kinds of individuals, though, looks set to far outweigh supply; it's predicted that by 2018 in the United States alone there will be a shortfall of between 140,000 and 190,000 data scientists compared with what's needed, alongside a shortage of a staggering 1.5 million managers and analysts capable of understanding big data analytics (source: McKinsey). Additionally, this skills shortage is cited as a key barrier to the growth of data analysis, with only 39% of organisations confirming that they have trained their analysts to cope with this new approach to business (source: KPMG).

One solution to the skills gap challenge is to embrace partnerships; by engaging with external suppliers highly experienced in the discipline as part of a long-term programme of work you can be confident in their expertise without having to commit potentially significant budget to put in place a dedicated in-house big data team or a regularly-repeated tendering and selection process.

A partnership offers numerous additional benefits, including the ‘first-mover’ advantage that can be gained by leveraging your supplier’s investment in continual innovation and optimisation, and the targeted guidance they can provide based on a better understanding of your overarching strategy and goals.

However, even if you do choose the partnership approach as a solution to your big data demands you will still need to have people in-house able to bridge the gap between your supplier and your organisation by providing specialist domain knowledge and reviewing progress regularly to ensure that your specific business requirements are being met. This may involve targeting your recruitment activities to focus on these roles, or updating your training processes to provide existing staff with the requisite skills.

It’s also important to make sure that people across your organisation are able to contribute to and benefit from your data by making it widely and easily accessible, encouraging utilisation and discussion and increasing the range of perspectives from which your data is assessed. This will also help embed the right culture into your organisation; a subject covered in more detail later.

6 Safeguard your long-term success

While the qualities that make data big provide great value for organisations in enabling them to gain a deeper understanding of their audience and environment, they also mean that the data is ever-changing, and as such on-going investment will likely be needed to ensure the long-term success of your efforts. For the greatest results this may even require fundamental changes to your organisation's underlying culture, including reviewing your strategic planning and marketing processes to take into account big data opportunities and risks, and optimising your front and back-office activities to ensure you work with up-to-date information and apply it appropriately to business decisions.

You may also consider adapting your project management processes so that a cycle of regular testing, review and refinement is put in place that ensures you're getting the best possible results from your data and enables your organisation to better react to changing conditions. For example the Agile project management methodology, which places emphasis on regular communication, flexible roadmaps and sprint-based delivery, can be applied to big data projects to enable models to be built up steadily as new questions are continually discovered and answered throughout the process.

To effect these changes, which may be fundamental to your organisation, you should ideally have a high-level sponsor who can ensure big data projects are given suitable business priority, and secure participation from disparate teams and departments. Big data touches many different roles and areas of an organisation, and while the owner of your activity will depend on the objectives of your strategy it's important to keep all necessary stakeholders involved and updated. Consider, too, the target metrics and KPIs you will put in place to enable

progress to be measured effectively; make sure these are tailored for each different role in your project but in all cases ensure they are well-defined and tracked back to your strategic goals.

As well as updates to your process a change in mindset may be required; for example big data can be applied to shake up your sales funnel by creating brand advocates and evangelists through the continued development of customer relationships after a purchase has been made. To enable these kinds of transformations it's vital that you communicate effectively with all members of your team to develop a shared understanding of the importance of big data, along with a central vision of what you hope to achieve. Of course, in all cases you should review progress regularly, refining targets and tasks as necessary to keep your projects on track.

7 Consider the legal implications

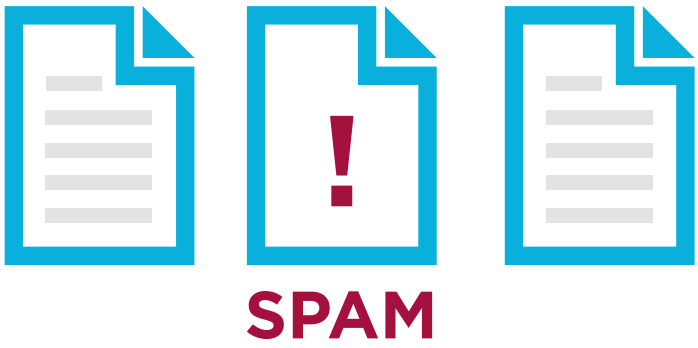
With big data often touted as offering seemingly limitless potential for organisations, it can be tempting to take advantage of every opportunity for analysing, manipulating and combining data sets to uncover new applications. However it is vitally important that any data you have access to is used appropriately; failing to do so could prove illegal, particularly in light of the tightening of privacy laws that has occurred in recent years. Close to 100 countries currently have data privacy laws in place (source: University of New South Wales), and the introduction of legal requirements such as the EU e-Privacy directive has significantly restricted how online user data is collected.

When creating your big data strategy then, make sure to thoroughly research any compliance regulations that may affect what data you're allowed to capture and how it can be applied, and build these requirements into your plan. There are a number of ways to help increase the security of your data, including implementing strictly-controlled firewalls as part of your infrastructure, setting up clearly-defined permissions determining who can access and edit data, and anonymising information by removing identifying details or aggregating results. You should also include in your metadata (mentioned earlier) any crucial compliance information, so that anyone who accesses your data can easily understand the situations in which it can be used.

Additionally, even if you are operating within the bounds of the law it is still important to ensure that your users' experience isn't negatively affected by your use of data. As such you must be transparent in all your communications with users as to what data you may be tracking and how you're going to use it. This is particularly important if information is going to be released beyond

your internal teams; for example, if public sector organisations wish to make data sets available for open source development projects or to enable external analysis to be performed.

Another key concern related to the security of externally-available sources is guaranteeing the accountability of any second- and third-party data you may have access to by building stringent checks into your selection process to ensure that your suppliers operate both ethically and legally. Additionally, if you have third-parties responsible for managing or analysing your data, make sure they are aware of any agreements you have in place and conduct their operations accordingly. Taking a partnership approach to working with your suppliers, as discussed previously, can further reinforce trust that they will follow any required guidelines, as the long-term, collaborative nature of the relationship helps instil a detailed understanding of your organisation, environment and industry that can improve their ability to make informed decisions as to what constitutes an appropriate use of data.



Finally, whatever data you use, remember to regularly review and cleanse it to ensure it remains relevant and recent. Not only is this important to enable you to get the greatest value from your analysis but it can be critically important in ensuring you remain within the boundaries of the law; for example, if changes to user preferences are not reflected in your data this could lead to communications being reported as spam and negatively impact your organisation's reputation.

Conclusion

Big data offers huge potential for increasing understanding of the business landscape and facilitating improved decision-making, but many organisations are still struggling to work out how to leverage it effectively. While this white paper does not have all the answers, it should have provided some questions to consider when deciding on your organisation's own use of big data, including:

- How might your technical infrastructure need to be updated to handle an increased volume, variety and velocity of data?
- What tools are available to manage big data, and which are most suited to your needs?
- What best practices should you be adhering to?
- How is your data split between first-, second- and third-party sources, and what are the consequences of this?
- What strategic objectives does your data need to help realise?
- Have all relevant team members contributed to and approved your strategy?
- Who will be responsible for managing your big data activity?
- Are your current processes suitable to support your big data requirements?
- Does your organisational culture enable people to easily extract value from big data?
- Are then any legal restrictions that apply to your use of data, and how will you ensure you follow these?

To find out more about how Box UK's consultants can help you get started with your big data strategy and ensure the appropriate systems are in place to deliver the greatest results, visit our **Software Consultancy** page or **get in touch** with a member of our team today.

Further reading

How P&G Uses Big Data To Turn Diapers Into Insights
BigData-Startups

What Is Big Data?
Forbes

Seizing the data opportunity: A strategy for UK data capability
HM Government

The top five ways to get started with big data
IBM

Enterprise Big Data
IDG

Going Beyond the Data – Achieving actionable insights from data & analytics
KPMG

Big Data Survey
MBN

Big data: The next frontier for innovation, competition, and productivity
Are you ready for the era of 'big data'?
Ten IT-enabled business trends for the decade ahead
McKinsey

Rise of the Datavores: how UK businesses can benefit from their data
Briefing: Inside the Datavores
Nesta

How many devices do you carry?
Sophos

The Growth and Management of Unstructured Data

Wikibon

box UK

Simply brilliant thinkers making software brilliantly simple

Strategy & Planning

User Experience & Design

Software Consultancy

Software Development

At Box UK our team of simply brilliant thinkers, consultants and application developers mastermind simply brilliant solutions to the world's toughest web and software projects.

It's why blue chip organisations around the world, from airlines to car manufacturers, international banks to global charities, choose to trust us with the execution of their digital strategies, from the development of enterprise-level custom software platforms to the delivery of award-winning multi-channel experiences.

So if you want an effective, compelling software solution, talk to us.

---

info@boxuk.com • +44 (0)20 7439 1900