Big data - Where do you start?

By Gary Allemann

Big data is currently a buzzword in the industry with many organisations pondering its value, virtues and implementing strategies to harness the benefits.

Countless organisations' understanding is that this data, when tapped into correctly, can provide a host of benefits including market intelligence and competitive advantage. However, one challenge many organisations still face is understanding what big data really is and where to start to take advantage of it.

However, there are still a number of myths surrounding big data and distinguishing myth from reality is the first step.

Some big data myths include:

- Big data is about external data: Forrester research finds that most organisations only analyse 12% of their internal data. Big data is about making more of the data you have. External data sources may be added to bring more insight, but this is not the main focus for many companies;
- Big data is about size: On average, an Enterprise Data Warehouse (EDW) houses about 15 Terabytes (TB) of data. Also, an average Hadoop installation increases from 150TB to 200TB. Therefore, big data is bigger than your existing data warehouse, however, only by an order of magnitude;
- Big data means in-memory (in-memory enables users to have immediate access to right information): Moving your existing Search and Query Language (SQL) processing to in-memory databases will reduce processing times; however this does not allow you to gain new insights. This data allows us to answer new questions without depending on structured schemas - whether in memory or not. Although in-memory is great for high-speed decision making on data sets, big data analytics are often batch in nature;
- We can use our existing technology: When relational databases entered the mainstream about 20 years ago, developers had to learn new skills, such as SQL. The big data revolution will also require us to learn new skills if we are to take full advantage of the benefits of unstructured data sets. SQL is not a long-term solution to exploit big data;
- Big data is about Hadoop: Although a range of technologies and platforms can be used both to support and implement big data analytics, Hadoop is the preferred platform of over 80% of big data implementations. Unlike the others, this myth is largely true - any serious big data analytics solution is likely to be based on Hadoop. Hadoop addresses key big data challenges allowing you to consolidate both structured and unstructured data quickly and provides cheap, scalable storage and analytics power;
- Big data is free: Although infrastructure costs are relatively low, the people costs are relatively high. Companies that reduce their dependency on big development teams - by exploiting easier-to-use platforms and supporting self-service - will get insights quicker, stay ahead of their competitors, and have a lower total cost over those that spend years on reinventing code themselves;
- Big data is an IT problem: Big data must have clear business goals to ensure intelligence is harnessed - this is not an IT problem, but rather a business issue. Big data is about finding the answers to critical questions about your customers, your channels, your markets and your products that can differentiate you from your competition, improve your customer's experience with you, and make you money;
- Big data is too complicated: Big data requires new skills and approaches. Yet, modern self-service data-discovery platforms, such as Datameer, exist, which shield you from the technical complexity and allows you to deliver quickly and easily;
- We need a data scientist: Most companies will be able to deliver big data analytics using a team of existing staff - business analysts, business managers and statisticians who collaborate, using a shared platform, to design and deploy appropriate analytics. The trick is to focus on the business problem, not complex technology;
- No more data quality problems: More data will typically mean more data-quality problems. Big data must be managed and

assessed for data quality just as any other data is - if accurate and trusted insight is to be gained; and

- Big data is just hype: Although big data is in its infancy in southern Africa, our colleagues and competitors in Europe and the US are delivering real value - in areas such as fault management, customer experience management, market-driven pricing, compliance projects, fraud management and much more. Big data is beyond hype. Companies that act first could build an unassailable advantage over their slower moving competitors.

Clearly defining and understanding the common big data myths will allow an organisation to fully understand where to start.

One of the first challenges the organisation will face, especially with regards to budgets, is whether to buy or build a solution. Building big data analytics using open source code could take a year or more; will require scarce (and expensive) skills and may allow nimbler rivals a critical head start. Packaged platforms can allow you to deliver insight in a matter of weeks, with minimal training and support for your existing people.

Big data is not as daunting as most assume it is, yet many organisations drag their feet due to the misconceptions and misunderstandings around big data. Big data can offer a wealth of information and value to an organisation, giving them competitive advantage and improving customer relationships. Opt for a solution that is already available such as Hadoop that offer quick implementation and provides real value in a short period of time.

ABOUT GARY ALLEMANN

MD of Master Data Management He is passionate about Information Communication Technology (ICT) and more specifically data quality, data management and data governance.
- Major trends impacting 2023 information and development planning - 27 Jan 2023
- Why data management is key to unlocking the digital transition of African banks - 12 Sep 2022
- Can data privacy concerns be managed when moving to the cloud? - 7 Sep 2022
- Effective WFH streamlined through data management as the next frontier - 8 Mar 2021
- Data governance is key to differentiating your customer experience - 20 Mar 2020

View my profile and articles...

For more, visit: https://www.bizcommunity.com