Bridging the big data divide

Part II – The discussions needed between IT and the business

July 2014

Your business is completely dependent on its electronic data. However, due to the way that the business has grown over time and the way that technology has changed continually, much of this data is now spread around in silos of storage with diverse capabilities. As the big data era takes hold, it is time to sit down with the business and sort out what IT can do to provide a fully supportive data platform for the future.
Executive Summary

Bridging the big data divide

Part II – The discussions needed between IT and the business

Effective data analytics will define the winners in the markets
Data is the lifeblood of any organisation. Through effective analysis and reporting, it can be built up through information to knowledge where business executives can then ensure that they are making the right decision in strategic areas. Poorly defined analytics, working against incomplete data sources and being incapable of dealing with mixed data types, will lead to partial information and knowledge being fed through – but with the business executive still believing that the right decision has been made.

Any decision on tools must be led by the business defining desired outcomes
The business is still in charge of its strategy. IT’s role is to ensure that the business has a technology platform in place that supports the strategy. Therefore, an understanding of what it is that the business sees itself doing in the near- and mid-term future is required so that IT can make sure it is addressing the right issues. IT has to become more of a partner to the business; it cannot remain as a separate department sitting on the edge of the business.

Discussions must be business led – the technology is secondary
IT has to move away from the use of technical terminology and jargon. The business is not interested in discussions around SQL v. noSQL databases, or in Hadoop as a persistent store. The key is to be able to talk the same language – how will what IT is proposing help the business meet its desired aims?

IT has to be the facilitator
With any problem, there will likely now be several possible options. It is not down to IT to choose one option directly itself, but more down to IT to examine what each option offers and present back to the business what these mean to it in terms of business cost, business risk and business value. The business can then make its choice on what solution fits best against these variables – and will then buy into the solution much better than if it had just directed IT to solve the issue.

Data analytics is not just a one-dimensional problem
Although some issues concerning data analytics will require real or near real time systems to be in place, much will be more asynchronous. Therefore, solutions chosen, where the underlying hardware is under the IT department’s control, can be sized against tiered storage and servers as to what mix of different analytics workloads are likely to be needed. Many workloads will be cyclical, with the need for resources required for these to be shared amongst less high-priority workloads when not in use. Conversely, when these cyclical loads are running, low-priority workloads can be put on hold so that any resources available can be used.

Building up discussions between the business and IT creates on-going benefits
The more that the business and IT enter into meaningful discussion, the more capable IT becomes in being a valuable advisor to the business. By sharing the output of the business analysis tools chosen, IT can continue to tune and nuance its advice to the business, working against the same data that the business itself is using to define its strategic and tactical responses to what is happening in its markets. IT becomes more a trusted partner, and the knock-on effects can have far-reaching positive consequences for other projects where IT can take a similar approach.

Conclusions
IT has to take a different approach in how it deals with the business. The growth of data volumes and types is already stressing how IT is supporting the business in its analytics needs. The continuing growth of BYOD and the imminent arrival of the internet of things (IoT) will only accelerate the growth in data volume that has to be dealt with. Only by entering into a full and authentic peer discussion with the business about what it really needs from its data analytics can the right systems be put in place. Get it right and everyone benefits. Get it wrong, and it is likely that the business could fail.
The rise of the data empire

Technology is now the foundation of the vast majority of organisations’ capabilities in competing on the global stage. The rapid rise in the power of available technology and the falling costs in server, storage, and network hardware, along with the global reach of the internet, has led to the ubiquity of technology within the day-to-day operations of any business.

If we were to liken this hardware to the skeleton of the business, then the applications could be seen as the muscle and organs. The data would then be the blood supply – dependent on the organs and muscles for being moved around, while also providing the very fuel that makes them work.

Although it is tempting to see the hardware (on which capital has been spent) and software (where licence money and maintenance are spent) as the main assets within a technology platform, it is the data that is the real business value.

Although it is tempting to see the hardware (on which capital has been spent) and software (where licence money and maintenance are spent) as the main assets within a technology platform, it is the data that is the real business value.

A server has a notional asset value – but this is rapidly diminishing from the point of order to the point of disposal just a few years later. A modern application often has no actual value – licences are moving to subscriptions; if you stop paying, you no longer have any access. However, as data volumes grow, the capability to extract more information and then knowledge from the pool of data becomes more apparent. Essentially, while the value of the physical environment tails off, the value of data grows. It is doubtful that a company’s main board will stare in awe at its datacentre and appreciate its beauty. However, IT could well be held accountable if there is a lack of capability to make decisions based on readily available, verifiable information covering as many areas as possible.

With different types of data being brought into the organisation on a regular basis – whether this be through images, voice, or video, or through different sources such as external information stores like Dun and Bradstreet or Lexis Nexis, or less structured feeds such as social network sentiment analysis through an external provider such as DataSift, or through web searches carried out by employees – it is now incumbent on the IT department to get to grips with data. Whether you want to regard this as just another data problem or lump it in with the ‘big data’ debate makes no difference – your organisation is more dependent than ever on making the most of the data available to it; it is down to IT to ensure that the right tools are in place to be able to provide the business with what it needs.

The growth in bring your own device (BYOD) and cloud computing is adding to the IT department’s issues when it comes to deciding on the optimal technology platform for the business. The incoming impact of the Internet of Things (IoT) will also have to be taken into account – just what new devices will have to be considered as data sources and providers, and just how will you need to deal with the new influx of data?

Data siloes have to be avoided; new data sources and types have to be embraced as they come along. Taking a data-centric view of the problem, based on discussing with the business what its desired outcomes are now and will be in the future, can lead to providing true differentiation in the markets.

With the right tools in place, IT can help in making data more fluid, ensuring that all knowledge workers and decision makers within the organisation have access to what they need, and creating an organisation where data is king. This then leads to an environment where decisions are made effectively and optimally, taking into account not only what lies directly within the organisation’s direct control, but also what is available across the value chain of suppliers and customers, and beyond that, to specific information sources and the public web.
Drowning in data

Just how much data do you have in your business? Start with the obvious sources – applications sitting on top of relational databases; not only how many applications you actually have, but also how many overall instances are there? Now, how about all those file stores with the documents, presentations, and spreadsheets of all your users? Are you logging customer calls coming in through your VoIP system and, if so, what do you plan on doing with these? The amount of data that is directly under your control is probably doubling every year, at least, and in many cases is growing at a faster rate than that.

Now look at where you believe your technology strategy will be going in the next few years. Cloud computing is probably already firmly on your radar; BYOD has been an unstoppable force. Will you be allowing the use of external file stores such as Dropbox, or trying to apply some measure of control through more enterprise-grade solutions such as Huddle, Citrix ShareFile or Accellion? What will be the mix of solutions based on applications and functions being sourced within your own owned or collocated datacentre, compared to how much will come from infrastructure, platform, or software as a service platforms? Where will the data underpinning the various parts of the platform reside – and how are you planning to deal with it all?

Much of the discussions here will be down to the IT department to sort between itself and its partners and suppliers. However, to fully support the business, you need to better understand what the business is trying to do and then put in place the right tools to ensure that the right data sources are made available – along with the right tools to make decisions, based on the best knowledge, to be taken.

Old style data analytics can no longer keep up with what is required. Just being able to analyse data in databases with hard-coded reports is now a constraint on the business.

The problem now is that the accessibility of data is too easy. If a company wanted to, it could view all its value chain as a data asset – as well as the whole of the internet.

Old style data analytics can no longer keep up with what is required. Just being able to analyse data in databases with hard-coded reports is now a constraint on the business. It generally only allows decisions to be made against what has already happened – increasingly, the need is to work against what is happening now and to extrapolate this effectively into the future. Different laws in different geographies, and the need to demonstrate areas such as information security to customers and other stakeholders, requires that reports be created demonstrating that governance is not just being paid lip-service to, but is actively being policed and maintained.

Without proper guidance, the business will either assume that only data within the organisation can be used, or that everything that has ever been created can be accessed and analysed in real time. They need help in order to understand what is feasible – and what is financially and securely possible. In other words – they need your help, but you will need to discuss matters with them in a way that makes sense to them: it is pointless going on about data warehouses, cubes, in-memory technologies, and Hadoop. Such discussions just lead to the glazing over of the eyes and the nodding of heads: not in understanding, but in a manner of “Whatever – just get on with it”.

The key is then to provide options to the business: ones that are messed against three main variables that they can empathise with. These are: the impact of the change on the risk profile of the organisation; the impact on cost; and
Bridging the big data divide

the added value that the option brings to the business. The business can then make an informed decision on what should be done – and IT can make it so.

Working to desired outcomes

The business is not interested in analysing everything – although it may seem to want to do so. It needs to have the right information to help it make decisions that will result in the right business outcomes. Without an understanding of what these desired outcomes are, IT will be working in the dark, and any chosen system will struggle to meet the business’ real needs.

Therefore, IT needs to sit down with the business and ask the right questions to find out what the desired outcomes are – and then make sure that the right data sources and means of blending and analysing them are in place.

The main strategic desired outcomes for organisations today will include optimising profit margins; increasing product or service sales; looking at launching new products or services; entering new markets; maybe growing through acquisition. Each of these has different requirements as to the types of underlying data that will be required and how everything will need analysing and reporting. None of them are nominally dependent on any specific technology – the business is not interested in that aspect. As IT, you will need to look at what knowledge will be needed to ensure that the move from the current situation to the future state can be carried out smoothly and effectively, and that any chosen systems will provide a long-term, flexible solution to meet and facilitate the business’ needs.

The business may already know what some of the data sources are that will be needed to make this happen, such as business information services from commercial sources. As IT, you may be in a position to identify better, cheaper sources, or additional information sources that can be used to add further value to those the business has identified.

It may be that there already are information sources that are already available but are not being used – for example, with silos of data created by standalone applications that only need integrating into the overall pool of data resources. Use your own sources to identify how missing data sources can be obtained and whether this should be done through implementing internal systems (i.e. buying in new applications to be implemented in your own datacentre or co-location facility) or through using external services (through subscription services or tapping into internet-based sources). At all points, make sure that the end result is not lost from view: all these data sources must be pulled together and analysed through a single system that can be used effectively by the business users themselves.

To ensure that the end result is met, it is important to bear the following areas in mind while discussing with the business or when assessing or acquiring systems:

- Capability to pull together different formal data sources across a wide range of internal and external platforms.
- Capability to include different data types, from existing databases to file stores (office and textual documents) and data streams (for example, machine data from production lines, building security systems) and, increasingly, to include image, voice, and video.
- Capability to blend these information sources in an intelligent manner so that analysis can be carried out across a broader set of available information, rather than just one silo.
- Capability for analysis to be carried out by a general worker using a simple interface, rather than being limited to skilled data analysts.
- Capability for results to be viewed by different people in different ways that suit their needs best, rather than using prescriptive and prescriptive pre-defined reports.

Therefore, IT needs to sit down with the business and ask the right questions to find out what the desired outcomes are – and then make sure that the right data sources and means of aggregating and analysing them are in place.
Bridging the big data divide

• Ability to use process flows to move analysis findings along a decision-making chain in an intelligent manner, building on previous findings as the process moves along.
• Ability to drill down through multiple layers of previous findings, along with the capability to add new data sources as required.
• Ability to include findings in other processes and to automate triggers, where applicable, so that findings can kick off actions as required.
• The ability to understand the underlying security of existing data sources, and maintain this, while being able to overlay extra security as and where required.

This list is the same whether you are approaching the problem from a business or a technology point of view. The only difference is then how the systems to make it happen are chosen – the business should not care: IT has to get it right.

Questions to ask the business

To ensure that your organisation does have an information management platform for the future, there are certain questions that IT has to ask the business.

<table>
<thead>
<tr>
<th>What data sources do you have access to—from the existing IT systems, departmental and personal subscriptions, and/or accessing free sources over the internet?</th>
<th>Only by identifying what is already in use can you start to pull everything together. Identifying all data sources gives you an idea of the types of data you will have to deal with, as well as the volumes.</th>
</tr>
</thead>
<tbody>
<tr>
<td>How rapidly do you need results for each desired outcome analysis?</td>
<td>Not all requirements need to be in real time. Running off a cyclical report on how well competition in the business’ market is doing can be a batch job where results being provided hours or even days late may not overly adversely impact the business’ capabilities to make the right decision. However, you will only know this by getting the business to truly define what the priorities are for each outcome analysis.</td>
</tr>
<tr>
<td>How many people need to create reports?</td>
<td>Analytics systems and subscription data sources will often be chargeable by the seat. Knowing how widespread the business believes it needs to give access to the full capabilities of an analytics platform will help you in costing out the options for the business.</td>
</tr>
<tr>
<td>Are all users to be seen as equal?</td>
<td>A user who needs to have access to create and run analyses against data sets is different to someone who needs access to the findings. For example, the business may say that 1,000 people need access to the system: what it may really mean is that 100 people need to be capable of analysing the data; 900 need to be able to consume that information in an intelligent manner in order to help in making decisions. Again, such knowledge can help in sizing and costing out solutions – and in planning the change management of any implementation project.</td>
</tr>
<tr>
<td>Are the skills for data analysis already in place?</td>
<td>The world is not full of ‘data scientists’; nor should it be so. Some aspects of data analysis will require certain skills, and it will be IT’s role, alongside HR, to put in place any training required to raise the levels of skills required within the user base. However, it is also IT’s role to ensure that any system chosen is intuitive enough so that any training can be kept to a minimum, and new users can be brought in to using the system rapidly and effectively.</td>
</tr>
</tbody>
</table>
## What will be the data flows within and outside the organisation?

Data and information security should be a high priority – what is being dealt with here is, after all, the organisation’s intellectual property. It needs to build up a full view of how process flows will operate: is this a purely hierarchical system within the organisation, or will suppliers and customers also be involved? Are other stakeholders, such as shareholders, part of the mix? How about reporting to central bodies, such as government departments or trade bodies? Again, only through understanding what the business will be doing here can you ensure that the chosen system has the granularity and capabilities of function to meet the security needs of the organisation.

## What output formats will be required?

The majority of analytic systems available can output reports in many ways. The most useful is as a ‘live’ report, where the user can drill down through the graphic to see how it was derived and to ensure that they agree with the underlying data sources. Bear in mind that some recipients may be outside of the organisation – and live reports must be accessible in a secure manner through the corporate firewall. However, not all reports can be provided in this way – some will still require hard copies: making sure that graphics can be exported for use in other applications such as Microsoft Office has to be a consideration.

## What impact will all of this have on storage requirements?

Big data analytics can lead to storage farms being required. However, modern systems should be able to minimise the amount of storage necessary through intelligent sorting – is this piece of data definitely required, possibly required, or not required at all? The use of Hadoop and its MapReduce capabilities will also help in sorting out what should be stored and what shouldn’t, and will also allow an optimised storage architecture to be employed to make the most of tiered storage systems. The blending of data sources should also avoid the duplication of data in persistent data warehouses, marts, or cubes: such artefacts should be temporary constructs used during analysis only, with the original data still being the main persistent stores.

## How long do you expect the proposed platform to work for the business?

Seeing beyond a short horizon is difficult. It is impossible to guarantee that any chosen technology system will still be fit for purpose in just a few years. However, make sure that what is chosen has the flexibility to embrace change – do not go down a route that makes it more than probable that a forklift upgrade will be required in a few years’ time. In essence – work to business cycles, not technology ones.
Presenting options back to the business

As discussed earlier, the business should not be interested in technology itself – only on how well or badly the technology deployed supports the business in its aims. Therefore, IT’s job is to present the options back to the business in terms it understands, bypassing all the discussions on the technical jargon and elegance of solutions.

The business is, in essence, only concerned with three areas – how much will this change cost; in what way does it impact the business’ risk profile; and what does it allow us to do?

Couching IT’s arguments in these terms will make it easier for the business to make the right decision, and also to buy in to the proposal chosen. Keeping the business involved in the implementation project through showing progress against these three areas will ensure that the project is a success.

| **Cost.** | Cost should not be the be all and end all of any discussion. It is obviously important but should only be one of the considerations. The overall concern has to be in how the cost balances against the other two variables of risk and value. For example, a ‘cheap’ solution that does not help the business could end up costing the business its livelihood by not providing the analytic support required. |
| **Risk.** | Risk is, increasingly, a major concern. As central bodies increase the fines for data breaches and brand value can be heavily impacted through poorly thought out information security, risk can often outweigh up-front cost as a concern. For example, enabling the business to monitor and analyse sentiment against its product brands can enable a rapid response to negative sentiment, preventing it from spreading and so lowering risk while optimising brand value. |
| **Value.** | The overall value to a business can be difficult to ascertain but, in essence, it comes down to how well the chosen solution will support the business’ strategy of bringing new products or services to market and in competing in its chosen markets. For example, by blending competitive information along with existing market figures for a similar product, the business will be able to make a much better informed decision as to whether to enter a new market with a product that will provide enough margin and sell in enough volume to make it worthwhile. |

Only when these three variables have been thoroughly examined and understood, and are then in balance, should a solution be chosen.

For IT, the options are becoming more complex. Should a chosen solution be acquired, implemented, and run in an existing data centre? Is now the time to look at a co-location facility? Can the functions required be obtained through a public cloud service? Each of these options needs to be taken into account when creating the messages back to the business. For example, is the current perception within the business that public cloud computing is too insecure? If so, can you pull together enough information to demonstrate that the perception is wrong, and show how a proposed provider, for example, has ISO 17799 and ISO 27001 certification? Arguments for or against each area have to refer back to what this means against the business’ own cost, risk, and value aspects.

Once the various options have been investigated and presented back to the business, it is then down to the business to make its choice based on the full knowledge of what it really needs. IT has to be available to answer any extra questions, to find out more information or to explain areas where the business is still unsure.
Once the choice has been made, IT then has to implement the chosen system as rapidly and effectively as possible, including all user training and change management. Such a project has to be carried out with as little negative impact on the business as possible: existing data may need to be integrated into the new system: use of blending and connectors, wherever possible, will minimise any need for data to be migrated, enabling a rapid switchover from existing analytics systems to the new one.

Conclusions

IT is being hit with a tsunami of data and of requests from the business for information to be extracted from such data. Without a new approach to data analytics, the business will not be fully supported, and competition will pull away in the business’ markets.

IT has to make sure that it truly understands what the business’ short- and mid-term strategies are, and to understand what the desired outcomes through the use of data sources are. IT also has to be able to identify what data sources are already in use and whether other ones can be easily obtained in-house, for free on the internet or through subscription services. It then has to look at what new analytics engines and approaches are available that can work with all the data types, volumes and sources of data that will be required, and make sure that whatever is proposed back to the business is couched in terms that the business can empathise with.

The key is for IT and the business to understand each other and to enter into any project as a single organisational team, rather than a command and fire approach from the business. By taking a more consultational approach to the issue of data analytics, the IT department will start to break down other barriers as well and place IT back where it should be – as a department that is there to facilitate the business.
About Pentaho

Pentaho is delivering the future of business analytics. Pentaho’s open source heritage drives its continued innovation in a modern, integrated, embeddable platform built for the future of analytics, including diverse and big data requirements. Powerful business analytics are made easy with Pentaho’s cost-effective suite for data access, visualisation, integration, analysis and mining. For a free evaluation, download Pentaho Business Analytics at www.pentaho.com/get-started.
About Quocirca

Quocirca is a primary research and analysis company specialising in the business impact of information technology and communications (ITC). With worldwide, native language reach, Quocirca provides in-depth insights into the views of buyers and influencers in large, mid-sized and small organisations. Its analyst team is made up of real-world practitioners with first-hand experience of ITC delivery who continuously research and track the industry and its real usage in the markets.

Through researching perceptions, Quocirca uncovers the real hurdles to technology adoption – the personal and political aspects of an organisation’s environment and the pressures of the need for demonstrable business value in any implementation. This capability to uncover and report back on the end-user perceptions in the market enables Quocirca to provide advice on the realities of technology adoption, not the promises.

Quocirca research is always pragmatic, business orientated and conducted in the context of the bigger picture. ITC has the ability to transform businesses and the processes that drive them, but often fails to do so. Quocirca’s mission is to help organisations improve their success rate in process enablement through better levels of understanding and the adoption of the correct technologies at the correct time.

Quocirca has a pro-active primary research programme, regularly surveying users, purchasers and resellers of ITC products and services on emerging, evolving and maturing technologies. Over time, Quocirca has built a picture of long term investment trends, providing invaluable information for the whole of the ITC community.

Quocirca works with global and local providers of ITC products and services to help them deliver on the promise that ITC holds for business. Quocirca’s clients include Oracle, IBM, CA, O2, T-Mobile, HP, Xerox, Ricoh and Symantec, along with other large and medium sized vendors, service providers and more specialist firms.

Details of Quocirca’s work and the services it offers can be found at http://www.quocirca.com

Disclaimer:
This report has been written independently by Quocirca Ltd. During the preparation of this report, Quocirca may have used a number of sources for the information and views provided. Although Quocirca has attempted wherever possible to validate the information received from each vendor, Quocirca cannot be held responsible for any errors in information received in this manner.

Although Quocirca has taken what steps it can to ensure that the information provided in this report is true and reflects real market conditions, Quocirca cannot take any responsibility for the ultimate reliability of the details presented. Therefore, Quocirca expressly disclaims all warranties and claims as to the validity of the data presented here, including any and all consequential losses incurred by any organisation or individual taking any action based on such data and advice.

All brand and product names are recognised and acknowledged as trademarks or service marks of their respective holders.