Bridging the big data divide

Part I – What your business needs to ask of its IT department

May 2014

We are entering a new era of the information economy. The advent of 'big data', where different types and sources of data need to be brought together and analysed rapidly and effectively, means that new approaches to business analytics are required. For an organisation to gain the most from big data, the business and IT functions need to understand exactly what is needed and what can be done. This report looks at what the business needs to be saying to IT so that IT can provide the business with the capabilities it needs.

Clive Longbottom
Quocirca Ltd
Tel : +44 118 948 3360
Email: Clive.Longbottom@Quocirca.com

Rob Bamforth
Quocirca Ltd
Tel: +44 7802 175796
Email: Rob.Bamforth@Quocirca.com

Bridging the big data divide

Part I – What your business needs to ask of its IT department

Dealing effectively with big data will be a defining point for winners and losers in all markets	Just being able to analyse formal data held in relational databases is no longer good enough. Businesses now need to be able to include different data sources of different types held both within and outside the organisation's control. The capability for a broad swathe of employees to blend and analyse multiple data sources in different formats, without the need to involve IT, will define who wins and who loses in the competitive markets.
Any decision on tools must be led by the business defining desired outcomes	The business is what counts; IT is there to facilitate what the business needs. Therefore, any choice of business analysis tools requires the business to define what it is it is trying to do – and then to talk to IT about the best way of meeting these needs. IT needs to have a good understanding of what the business' needs are right now and for the future so that a flexible business analysis platform can be put in place.
Discussions must be business led – the technology is secondary	The business and IT functions have tended to diverge in their ability to talk the same language. However, common discussion points need to be created to ensure that the right toolset is acquired and implemented. Jargon, technical discussions and comparisons between different architectures should be kept within the IT team itself.
The business needs to understand what is possible	Technology has moved on, and analysis of different data types across the value chain of suppliers and customers, as well as inside the organisation itself, is now possible. By asking the right questions of IT, the business can make sure that IT goes out into the market and finds a suitable toolset for the business' analytics needs: the key is in having the right discussions.
IT latency needs to be avoided	Where business reporting is based on IT providing report templates, any requirement for a new report involves time being spent by IT in creating the new report template. Meanwhile, the window of opportunity for the business may have passed. Ensuring that the end user has the capability to access an intuitive and flexible front-end system to analyse mixed data and to create meaningful reports, for both their own and other uses, will be of major benefit to any organisation.
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.

Conclusions

Effective business analytics, using mixed data sources, is of critical importance to today's businesses. Old-style reporting tools are no longer capable of providing the support a business requires; tools that can only work against data in single, formal stores are too constraining. An open approach is required that can take data from any source and allow a business person to carry out analysis in an intuitive and effective manner, without the active involvement in that process of the IT team. IT's role is to advise on what options are available; the business' role is to ensure that the options provided will offer the means to meet the business' desired outcomes.

quocirca

A data-driven world

Since the dawn of the computer age, technology has been harnessed to create, manipulate and understand data. From customer details to dealing with financial transactions; from supplier inventories to logistics tracking; computers have been used to both create and deal with more data than a human workforce would be capable of dealing with on their own.

However, whereas the main part of computer history has generally involved data that has a high degree of structure, we are now moving into a world where it is the increasing amount of less structured data that is becoming increasingly important.

The problem seems to be that the business is unaware of changes in the technical world that could provide it with the commercial differentiation to be more competitive in its markets.

Office documents; web searches; and social networking all provide inputs that can, and should, be used by an organisation to ensure that it has as much of the possible data available to hand in order to make the right decision at the right time – and in a format that makes sense to the person making that decision.

This also has to be seen against a distinct move in analytics capabilities. The old style 'business reporting' approach of running highly defined reports against data to see what had happened no longer supports a modern business. There is a need now to be able to look at what is happening in near real time – and also to look ahead with greater degrees of certainty as to what will happen in the future. The age of IT creating hard-coded business reports and the business having to go back to IT each time a new report is needed has to go: the winners in the new economy have to be able to analyse the data sources that are available to them rapidly and effectively across a far broader base of users and data types than previously.

The problem seems to be that the business is unaware of changes in the technical world that could provide it with the commercial differentiation to be more competitive in its markets. This short paper aims to provide the business person with the right questions it should ask of IT to ensure that IT and the business work well together to deal with the new information economy.

The modern information economy

Information is the most valuable asset an organisation has. Whereas it may have resided more in employee's heads in the past, increasingly, this information is being digitised and stored. However, without the correct tools in place to manage this information, and extract the vital knowledge that is hidden within it, organisations will still struggle to be competitive within their markets. When everything is held within a standard database (such as Oracle, Microsoft SQL Server or IBM DB2), standard tools can be used to analyse the data to provide standard reports. Today, even where such an approach is possible, it is unlikely to meet your business' real needs. Just what is a 'standard' report? What happens when a user wants something that is not 'standard'? Does IT have to be involved to create and test a new report – by which time, that crucial time window that the decision maker needed to hit with the knowledge the report would have provided them with may well have passed?

Information is the most valuable asset an organisation has. Whereas it may have resided more in employee's heads in the past, increasingly, this information is being digitised and stored.

Increasingly, though, what the business needs is the capability to analyse data that has been aggregated across a number of different sources. Again, this may be a number of different databases underpinning customer relationship management (CRM), enterprise resource planning (ERP) or other business systems. The speed of analysis and reporting has to be as near to real time as possible in order to allow the business to react to the dynamics of the market.

The ubiquity of the internet has opened up new data sources to the business. It is increasingly difficult to draw a line around any group of people or properties and state that 'this is the business' – the walls have come down and the need to deal with external data sources is becoming critical. The extended 'value chain' of your own business' suppliers and their suppliers, of your customers and their customers means that the more effectively you can deal with data from along this chain, the more streamlined your business processes can be and the more competitive you become.

It is increasingly difficult to draw a line around any group of people or properties and state that 'this is the business' – the walls have come down and the need to deal with external data sources is becoming critical.

This is further complicated as the traditional way that IT works has been eroded. Internal data centres are beginning to move to co-location centres and cloud computing is ushering in an era of distributed functions owned and run in different ways by different people. Some of these functions will be subject to commercial arrangements where data usage can be defined and managed; others will be free of charge, best-attempts services where you have little control over the format or even the quality of the data provided.

What about who should have the capability to analyse the data? Should this be down to a few centres of power? Should it be the new world of the data scientist? The aim should be to democratise the whole process, making data analysis available to everyone who needs to make a decision or feed into a decision within the organisation. Extending such access to others in the value chain in a secure manner would also add extra value in the majority of cases.

The modern information economy is increasingly diverse. Data has to be drawn in from multiple different sources in multiple different forms and blended in a manner that then helps a general worker to make a good decision or to feed filtered intelligence to someone else who will be making the decision. The data sources will not all be based on applications using formal databases as a means of storing structured data. Not all the data sources will be under the organisation's direct control. Many sources, such as social networks, will increasingly provide large streams of relatively unstructured data that need to be captured and filtered effectively before use. All data will have to be evaluated as to its security needs so that sharing along the value chain can be effectively managed so as to minimise risk while maximising value.

Your organisation is a player in the modern information economy – it is time to ensure that you understand the rules, and that you have the right tools in place to be a winner.

What the business needs

Your business needs to make sure that it has the right technical platform in place in order to deal with its present and future information needs. It is likely that your existing systems are not capable of meeting those needs; what is now needed is a review to see what can continue to be used and what needs to be brought in to provide a cost-effective information management platform for the future.

The first thing that the business must agree on is what the desired outcomes for business analytics actually are.

The first thing that the business must agree on is what the desired outcomes for business analytics actually are. This is likely to have changed from when older-style business reporting tools were implemented within the business. Look at where your business is now and where it needs to be in the future – 6 months, 1 year, 5 years out. It may be that the business is 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. Look at what knowledge will be needed to ensure that the move from now to then can be carried out smoothly and effectively. Look at what data sources will be needed to make this happen, and whether these are available already. Talk with IT as to how missing data sources can be obtained and whether this should be done through implementing internal systems or through using external services. Once the organisation has a true

view of what will be required, then it can sit down with IT and decide how to ensure that the right business analysis platform is in place.

A suitable information analysis platform has to include the following areas:

- 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 proscriptive and prescriptive pre-defined reports.
- 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.

Understanding your organisation's real needs allows for a better discussion with IT around what has to be in place and how to get to the position where business analytics becomes a business enabler, rather than business reporting being a business constraint.

Questions to ask IT

To ensure that your organisation does have an information management platform for the future, there are certain questions that you will need to ask of IT – and be willing to then discuss with them what will be required to meet the business' needs.

What capability do we have to deal with mixed data types?	If the focus within IT is purely or mainly around formal databases, then it will become necessary to invest in tools that will enable mixed data types to be included. This may include technologies such as Hadoop (an open source, scalable means of dealing with, and filtering, unstructured data) and NoSQL databases (a means of providing a base for managing and analysing unstructured data). Hadoop and NoSQL fall under the Open Source systems: there should be no direct licence to pay for the usage of these systems. IT can implement community versions of the systems to see how they work for them. However, once a decision has been made regarding using Hadoop and/or NoSQL, then it is likely that supported, subscription-based systems may be required to give the most commercially robust capabilities to the business.

Can we bring in external data?	The capabilities required to deal with external data sources is growing to an extent where just being able to manage the databases and files under the organisation's direct control is no longer good enough. Make sure that IT is considering the need to embrace external data.
What kinds of external data can we deal with?	Don't forget areas such as social networking: being able to pull in data from e.g. Twitter or other social networks can help an organisation to analyse just how people are feeling about the organisation and its competitors. The world of social networking is still dynamic – 'old' systems such as MySpace and Bebo have come and gone; Facebook, Twitter and LinkedIn continue to be relatively strong; new players such as Pinterest are showing promise. Being able to bring in and deal with streams of relatively unformatted data from a range of different sources will be required.
Can data be 'blended'?	A modern organisation must be able to deal with multiple different data sources at the same time. This can be done in various ways, most of which introduce significant delays into the analysis process. Ensure that IT are looking at tools that allow for data to be blended at source, creating a data pool that can be easily and rapidly analysed in near-real time to meet the business' needs. However, this blending should not require the user to involve IT. It should be possible to carry out the blending dynamically in an intuitive manner, with all security and audit requirements being met as required. Also, make sure that the chosen tools help the user in only showing data types that can be blended: not all data is capable of being used with other data types in a meaningful manner and the tools should be able to identify such issues and only present data types that can be blended. The tools should also be capable of ensuring the data is 'cleansed' to ensure that only accurate data is being used to prevent other data sources from having their value diluted.
What skills are required to use the analysis tools?	Ease of use has to be high on the list of buying priorities. A tool that is aimed only at highly data-literate people can only be used by those styling themselves as 'data scientists' – a dangerous position. The idea should be to enable anyone that has a need to analyse data to be able to do so. Interfaces should be intuitive; data visualisation should be effective and flexible so that different people can look at the same analysis in ways that make sense to them.
Can analysis be carried out collaboratively?	The capability for people to work together on analysing data allows other viewpoints to be brought into the mix and for better decisions to be made. Tools that work best for one user at a time make this difficult – a more sharable, centralised system, which is still capable of dealing with decentralised data sources, will be required.
How can the security of information be maintained?	Many existing databases will have their own security schemas associated with them, either via profiles being applied through the application or through profiles being applied to the database itself. However, once data starts to be blended across multiple different sources, other security issues may need to be considered. For example, does blending CRM data and transactional data lead to possible issues with personal identifiable data (PID) being moved around? Does opening up the value chain to access certain information assets lead to data protection issues, or threaten an organisation's certification, e.g. ISO 27001?

Can the analysis tools be used as a core part of the business' processes?	Self-contained business analytics have essentially reached the end of their useful life. Any analytics tools should now be capable of being integrated into other enterprise systems so that analysis can trigger actions to automate tasks and process flows. By streamlining these actions, organisations can find that their competitiveness is greatly improved.
How long do you expect the proposed platform to work for the business?	It is difficult for anyone to look further out than a couple of years in technology terms. However, IT should be able to state what parts of their decisions will be strategic (lasting for two years or more) and what parts are tactical (lasting for less than two years). This can help to gauge when the next cycle of analytics platform planning should start to take place.

What to expect back from IT

Based on the discussions, IT should be in a better position to go away to investigate, and then present, a range of options back to the business as to how to address the data requirement and analytics platform issue. IT may be looking at the various options including fully on-premise systems (ones that are installed and managed within your own data centre), fully outsourced systems (systems that are hosted and managed by an external supplier), or a hybrid mix of the two. Their arguments should be assessed against three variables – cost, risk and value.

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.
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.
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.

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

By tasking IT to come back with details couched in these parameters, the business is ensuring that IT will have to think about their options in the business' terms – not just on the pure technical aspects of any possible solution. This means that IT will come back to the business with options and arguments that give the pros and cons for each option in language that will give the business the information required to make the right choice on the tools that will best suit its needs.

Should IT not come back with options couched in business terms, make sure that you ask for clarification. In what way will the proposed option lower costs? Will this be only through up-front acquisition costs, or through on-going lower operation costs? Is any portion of it through softer savings such as better employee productivity and, if so, how can this be measured going forward? In what way will it lower risk? Is this through better audit capabilities, a better granularity of how information can be managed on a group role or an individual basis? How does the proposed option promote greater business value? Is this through a better capability to assess the greater market through including multiple external data sources, or through faster operations allowing more analyses to be carried out?

Ensure that the right information is provided – enter into meaningful discussions with IT and make sure that the team has the right information in their hands to be able to respond quickly and effectively to the business' needs.

Paytronix

Challenges

Paytronix, based in Newton, Massachusetts, USA, is a company focused on loyalty software for the restaurant market, serving over 200 restaurant chains in more than 8,000 locations.

Overall, Paytronix has several terabytes of restaurant chain customer data held within its databases, covering over 2 billion customer interactions. Paytronix wanted to offer its customers greater visibility and analysis of the information held within the data so that each customer could finely tune and optimise its own loyalty programme.

However, the data was not all held within a single data source, and covered a variety of different data types.

Paytronix needed to find a suitable extract, transform and load engine to create a new data warehouse to allow its customers to carry out their analyses in real time. However, the aim had to be to maximise the value of previous investments in the data platform while creating a more intuitive and powerful interface for their customers to their data.

Solution

Paytronix chose a full solution from Pentaho. By choosing Pentaho's Business Analytics Suite, Paytronix got both a data integration and analytics platform. The various tools provided by Pentaho allowed Paytronix to gain different advantages:

- ➢ Pentaho Data Integration provides the means of facilitating Paytronix Data Insights, providing the required ETL functions from multiple different data sources.
- ➢ Pentaho Business Analytics is used to identify patterns within the data to action discounts and make limited time offers available to the restaurants' customers, as well as identifying those who will buy without an offer having to be made.
- ➢ Pentaho's Dashboards, Analyzer and Mondrian are all used for data visualisation and to further help in identifying patterns and further opportunities for the restaurant chains.
- ➢ Pentaho Concierge services created an OEM strategy allowing Paytronix to embed Pentaho into its Data Insights system in less than 2 months.

Results

Through the use of Pentaho systems, Paytronix rolled out its Data Insights programme, offering three major functions to its customers:
- ➢ A deeper dive into customer data to identify actionable opportunities to drive customer visits and spend.
- ➢ The option to capture data from non-standard environments, including social media, to further bolster analysis.
- ➢ Visual interpretation of the available data in real time to uncover noteworthy trends.

Further, Paytronix found that the Pentaho solution led to an 80% reduction in ETL data processing times, as well as enabling non-standard data sources to be easily embraced. Paytronix's restaurant chain customers are now using the system to also identify loss-making and under-performing restaurants and use targeted loyalty offers to rapidly turn the fortunes of these outlets around.

Conclusions

The new information economy demands a new approach to how an organisation deals with its information assets. Just running reports against data held in formal databases will not enable an organisation to be effective in its markets. The requirement now is for a far more flexible approach – one that can include many different data types spread across multiple different sources while allowing the end user to be able to carry out their own analysis without the need to involve IT on a day-to-day basis.

However, IT needs to be involved upfront in identifying the right toolset that is going to meet the needs of the organisation; a toolset that will provide the capability to support the business' desired outcomes on an on-going basis.

By ensuring the business asks the right questions of IT in the right language, IT can make sure that it selects the right options of available toolsets. Through the use of a common language between the business and IT, the business can make sure that it chooses the right option, based on the advice from IT.

Business analytics will define who wins and who loses in any market: only through being able to bring together all the data that is required to help make the right business decisions can an organisation operate at the business speed it requires. It is imperative that, at each stage, the right decision is made – to do this the business has to make sure that it understands what is possible and how to best work with its IT team to make these possibilities a successful commercial reality.

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

quocirca