BUTLER
ANALYTICS

BIG DATA ANALYTICS

A Guide for Managers

February 2015

Contents

Other reports in this series:

Text Analytics – A Business Guide

Prescriptive Analytics – A Business Guide

Decision Oriented Business Process Management

Model Management and Governance

The production of these documents was supported by FICO

Why Big Data Analytics?

Big data is driven by the rapidly falling cost of acquiring, managing and processing ever increasing amounts of data. Clearly we shouldn't be accumulating more data simply because it is possible to do so, and the whole big data concept is predicated on the notion that we can derive value from our data assets. Some data has to be preserved for regulatory reasons, and all of it is fuel for analytics activities, so that businesses can make better informed and more effective decisions at all levels.

While we have mostly been impressed with big data infrastructure technologies, these are in reality not much more than plumbing – bigger tanks, fatter pipes, more of them, and different shapes. Commodity hardware and open source software have been the prime drivers for reduced data storage and processing costs. While traditional data management systems; relational database management systems (RDBMS) in the main, have handled transactional data very effectively, they are not so good at handling streaming data (financial market data streams for example), documents and text, images and other more esoteric data types such as geospatial data. RDBMS are not particularly good at handling queries and reporting needs either, and this has been mitigated to some extent by the emergence of data warehouse technologies, where data are formatted in such a way to make reporting activities more speedy and flexible.

As a result of these limitations a wide variety of new data storage techniques have emerged, and distributed database architectures have evolved to handle vastly greater volumes of data – the most well-known being the Hadoop ecosystem of technologies.

The concept of big data has become prominent because of the synchronicity of greatly enhanced technical capability and the reduced cost of data acquisition. This latter is driven largely by data generated through agents external to the organization, primarily customers and social data, and the rapidly evolving Internet of Things (IoT), where devices generate data at a very low cost (mobile and medical devices for example).

So the scene is set for the acquisition of massive amounts of data of various types, and for their processing at speeds which have previously been prohibitively expensive. This is a necessary but hardly sufficient scenario for using big data technologies. We need to add analytics to create value from an otherwise impotent resource. Analytical activities can be broadly divided into those which are concerned with historical and current activity, and those which predict future behaviors. Business intelligence (otherwise known as descriptive analytics) is a well-established set of methods and technologies which allows users to look through the rear-view mirror and analyze historical performance. Predictive analytics and prescriptive analytics allow businesses to learn from the past and predict future performance. The former uses data mining technologies to detect patterns in historical data which can be used in future activities, and prescriptive analytics determines the best use of resources given a set of constraints and objectives. These are complemented by business rules technologies and methods, which support the creation and maintenance of possibly thousands of rules which apply to operational activity within the business.

And so big data, as it has become known over the last few years, is largely an infrastructure issue. This infrastructure exists so we can manipulate and exploit low cost data resources, to drive high value analytical models, which improve both the efficiency and efficacy of business operations. To this end we need another layer of technologies, methods and skills to manage analytical activities, which typically produce thousands of decision enhancing, analytical models. And we should not forget the central roles of business know-how and analytical skills, since without them the whole edifice is of little value.

Such is the impact of big data analytics that it is useful to separate out the traditional use of information technology as a means of automating transactional activity and business processes, from the emerging use of technology as a means of automating business decisions – typically at the operational level, but also providing input to tactical and strategic decision making. Just as businesses differentiated themselves by the efficiencies they realized through transaction and process automation, so there is a new opportunity to differentiate through the efficiencies and improved effectiveness created through decision automation and management. Big data analytics is the hub around which much of this new opportunity will revolve.

What Is Big Analytics?

Big data has been defined as data which occurs in volumes, varieties and velocities which cannot be handled by traditional database technologies. To this end a large number of database management systems, distributed database architectures and associated utilities have emerged which handle these requirements in a manner that is affordable. And yet handling the data is only a prerequisite for the real task of creating value. In order for this to be achieved it is necessary to have a stack of analytics tools which not only exploit big data, but allows users to explore it in a manner that hides the considerable complexities. Indeed complexity is the necessary price paid for handling massive volumes of greater diversity of data.

The four traditional methods of deriving value from data apply equally to big data, and in this sense there is nothing new here. These methods attract different applications:

- Business Intelligence has been democratized during recent years, with business users gaining direct access to the data and tools which support their own needs. Big data should not inhibit this process, and so tools are needed which provide not only business users, but also analysts and data scientists with a broad range of capabilities to explore data and discover relationships within it. Hiding complexity is a key requirement here.
- Business rule management is concerned with the application of rules to operational activity, and these rules are often numbered in the thousands in many organizations. Since these rule are not only manually created but result from activities such as data mining, their number can increase dramatically, and so a robust, high volume business rules management system is essential.
- Predictive analytics, which use data mining techniques to discover patterns in historical data, represents a core value creation method. With the emergence of big data we are witnessing the creation of many more predictive models, and greater rigor in needed to manage these models in an effective manner. This is required by regulators in many

cases, but is also needed to ensure the accuracy, easy modification, and adequate monitoring and documentation of models. A failure to do this results in predictive model chaos, presenting real dangers to the organization.

- Prescriptive analytics uses optimization techniques to find the best deployment of resources based on a set of constraints and objectives. Big data broadens the use of these techniques considerably, supporting more complex optimization problems, which in turn drive greater efficiency.
- Decision management is an overarching infrastructure for solutions which accelerate the cycle from analyzing data to delivering operational decisions that improve efficiency and efficacy. It gives business experts greater control to manage and improve business strategies through the use of business rules, predictive modeling and optimization technologies.

Big data brings with it big analytics – otherwise the data will simply be underutilized. The analytics in turn enable the automation of decisions, and this is the new territory that is up for grabs in all businesses. Managers are becoming increasingly aware that efficiencies and efficacy are no longer simply the province of transaction and process automation, and that decision automation can deliver very high returns on investment - in some cases an order of magnitude greater than traditional systems and applications investments. Just as transaction and process automation require an integrated infrastructure, so do decision automation systems, and this will become a more pressing issue with the adoption of big data infrastructure, and the proliferation of decision automation it will enable. Decision management provides the infrastructure, methods and disciplines required to address this.

Methods in Big Data Analytics

On the face of it big data analytics simply requires data, analytics tools and some means of deploying analytical models into the operational environment. Such a simple minded approach would be a sure route to ruin, since there are many skills, resources and methods which need to work together if big analytics are to deliver their promise.

Firstly there are considerations of the data itself. Big data is no different from any other data in many respects. It needs to be understood and pre-processed before it can be used in analytics activities. This means that business users, analysts and data scientists must have the tools which support data exploration and visualization, while hiding the complexities associated with big data. Understanding the meaning behind the data is absolutely crucial if we are to avoid deploying meaningless predictive models which are no more than 'accidents' of the data. However big data does present some unique challenges, and particularly data which are characterized by many attributes. There is a common misconception that large volumes of data cannot lie and mislead. They can, and they do it in ways which require skill and experience to detect. While big data may be relatively new, there is no substitute for experience.

There is now a plethora of analytics tools, and particularly predictive analytics, which provide stand-alone model development environments. For modest analytical activities these may be sufficient, but for enterprise needs they will inevitably be deficient in management, integration and deployment capabilities. Hundreds and possibly thousands of predictive models are typically the norm in organizations which automate their decision processes. Documentation, visibility, monitoring and maintenance of these models requires a decision management infrastructure. The alternative is a rapidly decaying ability to manage the decisions that predictive models are making. Model building also needs to be integrated with model validation, monitoring and modification, and so an integrated design, build, validation and monitoring environment is needed. Anything less and trouble will surely follow.

The reality of most predictive models is that they will be deployed into mainstream production systems – sales, finance, marketing, purchasing and most other activities relevant to a given organization. Since the processes associated with these activities are often created, deployed and monitored using business process management (BPM) techniques and infrastructure, it makes sense to integrate decision automating predictive models with the BPM environment. To this end a recently ratified standard known as Decision Model and Notation (DMN) has emerged to link mainstream business processes with the decision models they use. This is absolutely essential if widespread adoption of decision enhancing technologies is to be successful. Anything less and we end up with the islands of information and automation that plagued business applications prior to the adoption of integrated application suites (ERP and CRM for example).

There are also different mechanisms for deployment which must be considered. Many knowledge models can be represented in a manner (typically decision tables and trees) which are easily deployed in a business rules management system (BRMS). This has many advantages, and not least the visibility to business users, ease of modification and all the benefits associated with a central repository. Other knowledge models cannot be represented in this format, and so it is crucial that the decision management environment supports a unified view of all decision models deployed and under development within the organization.

Finally, not all models are predictive in nature. Some are needed to optimize the use of resources, and it is inevitable that other methods of optimizing and predicting outcomes will emerge. These can often be used together and this is only feasible in an integrated environment.

Despite the hype, big data is very similar to 'old data' in many ways, although it does present some new challenges. We shouldn't be carried away by the notion that more data is necessarily better data – only someone who really understands the business domain can say whether that is the case. However the availability of large volumes of diverse, low cost data does present many new analytical opportunities to create optimized decision models. Many businesses in diverse industries are realizing the benefits of big data analytics, and those which are most successful are the ones who have chosen to take an integrated approach based on decision management methods, technologies, skills and disciplines.

Big Analytics Applications

Big data supports a wide variety of analytics activities, but in the business environment it is predominantly interactions with customers where the technology is applied. These applications can be as diverse as estimating the likelihood that a patient is readmitted to hospital, to the likelihood that customers in a supermarket might buy milk when they buy bread. The most mature environment for 'big data' is to be found in the financial services industries, where activities such as fraud detection, customer defection, and delinquency have benefited from predictive models and business rules management for many years.

The swelling volumes of big data are mostly customer focused. The data generated within the business does not grow so quickly, it does after all require people to create it. However data that are created by the customers themselves is exploding in volume – social data, web site data, data from devices, and various other mechanisms for collecting low cost, diverse data from customers. While we tend to focus on data volumes, and the term big data encourages this focus, it is the diversity of data that tends to add most value in customer focused big analytics. This means we have to embrace data other than the familiar record oriented data held in transaction databases. Text data which might contain customer feedback, interviews with call center operatives, social data, comments on web sites, and other sources, adds a dimension that cannot be gained from traditional sources. Text analytics can be used to assess sentiment so that early detection of defection, fraud and other behaviors can be spotted with more accuracy.

With the emerging Internet of Things (IoT) we are increasingly able to process very high volume streams of data, which often need a real-time response. To this end stream processing is becoming a major form of big analytics. Typical applications include responding to customers in real-time as they move around a web site on their mobile (or other) devices. This might also include the use of location data – another data type that is well accommodated by big data technologies and analytical methods. As the scope of the IoT broadens, many businesses will need to apply analytics to real-time data streams, and it is here particularly that a mature decision management environment will be absolutely essential. Real-time is much less accommodating of mistakes and errors, unlike batch processing, which is the dominant use of big data technologies at the present time.

Increasing numbers of decision automating, predictive models will be applied to real-time scenarios. Optimizing decisions in the 'now' will increasingly become the determiner of success or failure. The applications range from monitoring complex processing plant, where tens of thousands of sensors might be installed, through to real-time processing of credit card usage to detect fraud. All industries will be affected – healthcare, manufacturing, telecoms, financial services, retail and most others.

And so the complexity and scope of big analytics will grow considerably over the coming decade. We have seen many, many times that complexity is the greatest threat to the successful use of information technologies. Big analytics cannot be treated as a boutique application that bolts on to existing applications. It needs its own methods, infrastructure and tools, and we predict that it will very soon become the tail that is wagging the dog. To this end it cannot be stressed enough that an extensive, mature, decision management infrastructure needs to be installed so that decision models can be monitored, modified, developed, verified and deployed in a manner that is safe, and where management retain control of the automated decisions that big analytics enable.

Big Analytics Case Studies

From looking at various successful case studies where big analytics has been employed, it is clear that success depends on a good understanding and definition of the business problem, and the ability to embrace a wide variety of relevant data sources, which might be quite diverse in nature. Predictive models then need to be managed through their own cycle so that performance can be monitored and changes made when necessary. These case studies are courtesy of FICO, the sponsors of this report.

- A leading North American supermarket chain wanted to target and interact with customers earlier in the purchase cycle. To achieve this it would be necessary to increase the depth and speed of analytics. Time-to-event (TTE) models were already used that pinpoint when a customer is likely to purchase particular products – often resulting in a lift of 150% in average visits for redeemers over nonredeemers. The business engaged FICO to see whether a combination of descriptive and predictive analytics could improve matters further. The descriptive technique used established similarity based customer groupings (called “archetypes”). This allows the retailer to reach out to customers with greater accuracy, and significant improvements in customer interactions have been realized.

- As the telecom industry increasingly focuses on precise management of risk and reward, one global player in this industry determined to address the marketing efforts which resulted in problematical customers. It has used big analytics to move beyond traditional credit classes to more granular analytic segmentation that separates populations by credit risk and customer lifetime value. The big analytics initiative uses a wide range of data to not only predict customer behavior, but also balance all the key elements of risk and reward in its decision to prescribe the best action for maximizing discounted cash flow over time. This has resulted in a significant fall in valued customer attrition rates, knowing when to apply leniency and when to act.

- A large North American bank employed big data analytics to address several issues – new product development, strategic segment growth and customer retention. Big analytics was used to help develop a new credit card for a particular target segment. This involved rigorous research, strong analytics and alignment of sales and marketing plans. Advisory services for affluent customers were improved by embracing a much broader data set and applying analytics. Customer retention was significantly improved after the bank engaged FICO to help with analytics. This involved analysis of thousands of characteristics and the creation of predictive models which resulted in much greater take-up of new offers. The primary lessons learned were that disparate data sources create accurate predictive attrition models; that timeliness is very important, and that large data volumes were critical to the effectiveness of analytics.

- A manufacturer of autos wanted to improve sales of one of its models. To achieve this a mobile marketing campaign was to be launched with high impact messages.

> This business decided to use FICO's Big Data Analyzer to gain insights and build models from a broad range of data sources including Twitter, and to employ text analytics. Sentiment analysis was carried out with extraction of high frequency product differentiators. This resulted in a highly successful mobile campaign.

What is most noticeable in these and many other case studies is the diversity of business problem and the uniqueness of the solutions. It is also clear that in-depth understanding of the business domain is essential, and to this end the use of data exploration and visualization techniques is essential. The multi-discipline nature of building and using predictive models means that an integrated decision management environment is crucial.

Big Analytics Strategy

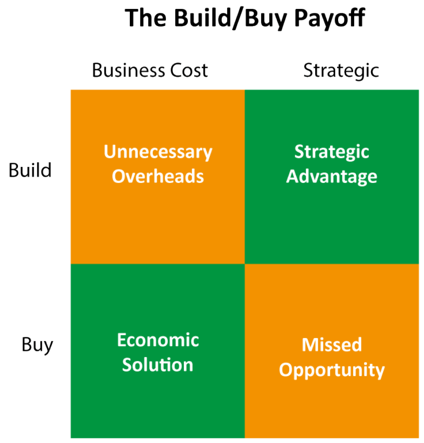
The starting point in a big analytics strategy has to be business need. This means that management has identified areas of business activity that might benefit from more accurate and efficient decisions. As such there is a need to coordinate investments so that these aims are achieved, with efforts to acquire relevant data, the necessary infrastructure and skills, and an enterprise wide decision management capability.

Big analytics need big data, which in turn needs low cost data. Very few organizations will willingly increase their data costs unless there are very well defined payoffs, and even then it is always best to get someone else to create your data for you. In the main this means customers through self-service mechanisms, and other sources such as social data. And the efforts to acquire low cost data should embrace diversity whenever possible, since this has been shown over and over again to be the key to better predictive models.

With a data acquisition strategy in place we can consider the infrastructure that is required to manage, store and process it. This is specifically the domain of big data and technology platforms such as that provided by Hadoop. This is an infrastructure issue, since big data alone is simply an additional cost which is borne in the belief that the data will yield some value.

Big data infrastructure is the necessary condition for big analytics. This is where we start to uncover value, and find more efficient and effective ways of dealing with customers, supplier and even employees. While big data is largely the domain of technicians and specialists, big analytics touches most parts of an organization – business managers, IT, data scientists, analysts, and of course the customer. Because of this it really is not sufficient to see big analytics as something that can be hived off to a team of analysts and data scientists, with the expectation that predictive models will appear which can then be deployed in a production environment. Regulatory requirements, performance monitoring, domain expertise, business process design and a raft of other issues need to be accommodated in the predictive model lifecycle. To this end an integrated decision management environment is essential.

Whether an organization should build or buy its big analytics capability depends on the nature of the application, as shown in the diagram below. This is a simple payoff matrix which shows that if an application is simply considered a cost of doing business, then it is probably



best to buy, since there is very little room for a strategic advantage. However when predictive models can make a large strategic difference, then those opportunities should be grasped and bespoke solutions built. Of course it isn't completely black and white, and a mixture of build and buy can be used when appropriate.

There are also questions which should be answered concerning the necessary infrastructure and tools. Some businesses might decide that they do not want the cost of supporting a big data and analytics infrastructure, and opt instead for a cloud based option. Others, typically larger organizations might opt for in-house capabilities or a mixture of both.

The use of big analytics does require some strategic analysis. Businesses that are happy to simply follow the herd where information systems are concerned, might have to accept the added responsibility of doing something a little bit different at times – simply to realize the benefits they anticipate. A quarter of organizations surveyed expect big analytics to deliver some form of competitive advantage. This is to be compared with only ten per cent who expect traditional process automation to deliver an advantage (Butler Analytics Survey). And so expectations are higher – and so they should be. Organizations which are prepared to grasp this particular bull by the horns often gain a significant advantage over their competitors.

About Butler Analytics

Butler Analytics was founded by Martin Butler, recognized as one of the preeminent technology analysts in Europe. He is best known as founder of Butler Group which, prior to its sale in 2005, was the largest indigenous technology analyst firm in Europe.

The focus for Butler Analytics is decision automation and management technologies, methods, and strategies, since these represent the next major wave in the automation of business activities, with associated improvements in operational efficacy and efficiency. The technologies covered include predictive analytics, big data analytics, optimization, business rules management and business intelligence.