The Bloor Group

New Directions in Analytics

How a Modern Rapid Analytics Platform Enables the Business

Robin Bloor, Ph.D.
& Rebecca Jozwiak

WHITE PAPER

Executive Summary

This executive summary touches on most of the business and technology topics covered in this paper, providing our conclusions as a short series of bullet points.

- Until recently, analytics was a relatively niche business application area used by large organizations such as insurance companies, pharmaceutical companies, government departments and so on. Disruptive technology developments changed this, primarily the advent of inexpensive parallel hardware configurations and associated open source software, which dramatically reduced the costs and heralded the IT revolution that goes by the name of Big Data.
- Businesses from many industry sectors became interested in exploring, adopting and exploiting the technology. Many Web 2.0 businesses were founded, and they succeeded by exploiting the disruptive technology.
- However, there was and still is an analytics skills shortage. Indeed, for many companies it is the major barrier to exploiting the Big Data revolution.
- Exaptive is an intelligently designed and purpose-built platform for developing and implementing every variety of analytics applications (embracing both cognitive and predictive analytics). We see its virtues as follows:
    - It is Web-based and easily deployed.
    - At the interface level, it is purpose-built for data science activity and is consequently easy to learn and easy to use.
    - It facilitates data integration, providing for genuine data self-service.
    - It accommodates all staff involved in the data science activity: the data scientist, the business analyst, the line manager, the end user and the software developer.
    - It enables the individual data scientist, the data science team and the company to build and implement analytical applications and gradually expand both their skills and the analytics toolbox.
    - It provides a marketplace, the Xap Store, that can be explored for new capabilities and new application ideas.
    - Above all else, it provides a versatile rapid development environment.
    - Given that data scientist skills are in short supply and expensive, the fact that Exaptive enables rapid development, significantly shortens the discovery process and improves data scientist productivity leads to faster time to value, which may be its most important feature.
- In our view, businesses that are looking to expand or establish a Big Data analytics strategy would be well advised to consider its capabilities.

The New World of Analytics

The Hollerith Tabulating System, invented by Herman Hollerith, was an electromechanical device that counted data recorded on punched cards. It was the first such statistical application of a computing device and was, as history records, used by the United States Census Bureau to process the data gathered in the 1890 census. The data was undoubtedly "big data" given the data volumes and the technology of the era – and the application was analytics.

Today's technology has moved light years beyond that. We've had to invent new words like "zettabyte" and "yottabyte" to describe vast volumes of data we now store and process. But the most compelling applications we run are still analytics.

There have always been analytics applications that were "edge applications," ones that tested and pushed back the boundaries of computing. Whether it was breaking German encryption codes in the Second World War or the modern day processing of the endless petabytes of data recording the subatomic collisions in the Large Hadron Collider (LHC), the application is analytics. For such edge applications, abnormal computing power is required. The breaking of the German codes required the invention of new computing devices and the LHC utilizes the largest computing grid in the world, with many thousands of servers.

The Genesis of Commercial Analytics

The use of expensive technology follows an easily understood pattern. It is used first by those with the need for it and the large budgets to pay for it. So the pioneers are usually wealthy governments and, if the technology has military application, the first customers will be the military. If the technology has application in businesses, then the largest of businesses will be first to adopt. From there, as it becomes less expensive, the technology gradually works its way down to the small business and finally, if it is appropriate, to the consumer.

Technology that is born inexpensive follows a different path, dashing into mass markets at break-neck speed in pursuit of volume sales. But analytics is not by any means a consumer application. It is an enabling technology, an ensemble of tools and capabilities, whose sole purpose is to extract useful knowledge from data and put it into action.

The major attraction for the early adopters of commercial computing – governments, banks, insurance and pharmaceutical companies – was the ability to run statistical analytics on the data they gathered. The professions that exploited the early analytics capabilities were those of actuary and statistician, and the applications they built were costly by any reasonable metric. The computer hardware was expensive, the software was expensive, the IT support from developers and operational staff was expensive and the professionals that used the applications were highly paid.

Consequently, in the years that followed the mainframe era, data analytics did not proliferate widely. It was adopted by large companies in other sectors – retail, utilities, manufacturing, transport and so on – but did not expand much further. By comparison, business intelligence (BI) software, which had its origin in spreadsheets and reporting tools, was far more widely adopted, gradually becoming a common feature of the business environment.

Feedback on business processes was thus provided by dashboards, data visualizations, drill-down capabilities and performance monitoring software with its KPIs. It was not difficult to train staff to make productive use of such software and embed such capabilities into the way the business worked. But such software was very limited in the benefits it conferred.

The Perfect Storm

Analytics marched into the Big Data limelight because its costs dropped like a stone. In 2004, Intel became obliged, as it brushed up against the technical limitations of silicon, to add extra processors to its x86 CPU chips and add more as the years rolled forward. The chip industry went multi-core, which in turn provoked the software industry to develop parallel software to exploit those multi-core chips. And then, straight out of left field, came Hadoop, and with it, a whole parallel software ecosystem evolved, driven by Apache open source projects that included libraries of analytics capabilities and a host of new NoSQL databases. And while this almost-free software revolution was in its infancy, Amazon introduced a further disruptive factor by launching its cloud service.

By around 2011, the beat of the Big Data drum could be heard everywhere, and the world of analytics was reborn. A perfect storm of hardware and software developments made it possible. Spinning disk was gradually dying, deprived of oxygen by solid state disk (SSD) that was faster, more reliable and sped up in harmony with Moore's Law. RAM was falling in price, and every new server sported more of it. RAM, 100,000 times faster than spinning disk, had rarely been exploited by parallel software, but now it became a vehicle for it, particularly with the advent of Spark, a Johnny-come-lately open source platform with its own analytics libraries. Hadoop, with its scale-out architecture, could unite the power of thousands of servers and exabytes of disk or SSD, and Spark could exploit petabytes of memory, as long as you could bear the hardware cost.

If nothing else, parallelism meant that you could accelerate analytics dramatically as long as the underlying software ran in parallel. This wasn't a Moore's Law level of acceleration (a factor of two every 18 months), it was an unprecedented level of acceleration: a thousand times or more, at a relatively low cost. So we entered a new world. We could process volumes of data we had never dreamed of a few years earlier, and we could do so at unprecedented speeds. Analytics was no longer the same application.

The Birth of Data Science

In 2010, the profession that once bore the name statistician or data analyst acquired the sparkling new title of data scientist. If there was any justification for this revision of the English language beyond the machinations of marketers, it was that the job of the statistician/data analyst had been fundamentally transformed by the dramatic improvement in the capability of the underlying technology.

There were no new mathematical techniques or tools that suddenly appeared and no methodological changes in how such techniques and tools should be applied. The speed of analytics and related applications accelerated dramatically and this changed everything else:

- Much larger volumes of data could be analyzed.
- Many new data sources could be added, including sources of "unstructured" data.
- The use of machine learning algorithms became practical and thus desirable.
- Streaming analytics was far more practical.
- The cost of analytics reached a point where many companies could adopt it.
- The analytics software environment was fertile and included many open source capabilities that could be adopted at very low cost.

- The cloud made the speed at which an analytics capability could be built and used both fast and inexpensive.

The critical limitation to the business world's ability to exploit data science was the number of people with the necessary skills: the data scientists. Nobody predicted the spurt in the demand for such skills, so the universities and colleges were unable to provide a readily educated stream of graduates.

The Scarcity of Big Data Skills

Currently, there is a 50% year-on-year growth in the advertised vacancies for data scientist positions, and the combined salaries of all those vacancies (roughly 200,000 jobs) is estimated at roughly $20 billion. It has been observed that most data scientists could get a 20% pay rise simply by applying for a new job. College graduates with relevant degrees are in short supply and command high starting salaries. McKinsey has projected that "by 2018, the US alone may face a 50 percent to 60 percent gap between supply and requisite demand of deep analytic talent," and Gartner projects a 100,000+ person analytic talent shortage through 2020.

The reality is more complex than the headline statistics suggest. Organizations with well-established analytics operations can now achieve much more at far lower cost. They may require more staff but can continue to hire graduates and train them. For most new web businesses, social media companies and gaming companies, analytics is the beating heart of the business. They usually have a data science capability from the onset, along with technical Big Data skills. They can also probably train graduates, and as they are fashionable companies to work for, they have few recruiting problems. The major consultancies and systems integrators see analytics as a major opportunity. They can bid up salaries and thus have few recruiting problems as well.

This puts the businesses that wish adopt analytics for the first time in a difficult situation. Most of the associated skilled jobs – data scientist, data mining engineer, machine learning engineer, data architect, BI architect, Hadoop engineer, senior data scientist and so on – command six figure salaries. And the company itself will have little idea of how to define and establish an effective analytical business process within the organization. They will probably need the assistance of a consultancy to do so.

Organizational Culture and the Analytics Process

While analytics is a broad field of activity, it can be roughly categorized between predictive analytics (sophisticated decision support) and the multifaceted analysis of business data (exploration and discovery). The first category corresponds to useful feedback about business activity and can be taken to include all BI applications (reporting, dashboards, performance monitoring, OLAP and so on) as well as predictive analytics.

The second category of analytics is a natural part of the R & D process of the business: having the possibility to impact business strategy, tactics and even organizational structure. The conundrum for businesses new to analytics is how to staff an analytics department, how to accommodate the analytics business process within the organization and how much budget to allocate to it. Even with consultancy assistance, organizations unfamiliar with the analytics business process and the possibilities of the technology may stumble.

Aside from employing the right staff with the right skills, choosing the appropriate tools – the IT components – will be key to success. As we described, the technology of analytics has been transformed, and it is affordable even for companies with a relatively modest budget. The software can be inexpensive (consisting primarily of Hadoop and elements of its open

source ecosystem) running on powerful inexpensive hardware (commodity servers running in parallel and possibly in the cloud) and employing selected components from the wealth of new analytics tools and platforms.

However, the choice of analytics tools and the underlying platform will be a critical element in defining and enabling the analytics business process. It is precisely here where the rubber meets the road.

Exaptive: An Analytics Workshop

Historically, software development products tread a natural path of evolution. Initially, point products emerge that deliver useful development capability, and then heterogenous tool sets develop that provide broad coverage of the application area. Next, you see integrated platforms that thread together the complementary components in a well-integrated way. Finally, we tend to see rapid application development introduced. We witnessed such software evolutions following the birth of relational database in the 1980s and again with the advent of data warehouses and BI. We see it playing out now with analytics.

In recent years, we have seen a variety of new analytics products emerge to complement or compete with traditional analytics tools such as SPSS and SAS. Exaptive is the first we have encountered which genuinely moves the needle. It is one of those software products where the demo provides you with an accurate impression of how the software behaves in practice.

Before we describe that, it may help to outline some of the fundamental aspects of the product. Exaptive Studio is a browser-based integrated development environment (IDE) that provides a single coherent interface for developing data-driven applications or whole workflows.

Consider a data science project. A particular objective or area of focus is decided upon, then possible data sources are examined to determine how to gather the data that is to be analyzed. Depending on the sources, it may be necessary to transform the data while gathering, and once collected, it will be examined and quite likely cleansed. Following that, various analyses will be performed, and depending on the results, the whole process or individual parts of it may be reiterated for the sake of refinement.

If the ultimate outcome is usable knowledge that can be made available to staff to improve decision making, or even delivered directly to applications to automate particular decisions, then the analytic processing will need to be implemented as a regular application that integrates with the relevant systems.

Exaptive Studio's IDE combines data flow with an object-oriented/service-oriented approach to software development, making it possible to develop applications and workflows – called Xaps – in a visual drag-and-drop manner. Extensive reuse of software components or whole programs can be achieved with little need to dip into program code.

The way the visual interface works is illustrated by Figure 1, which shows a screen shot of the creation of an application to search public medical information. Associated program components are linked together in groups. For example, the figure shows the ***Write to SQL Database*** grouping with two of its components, *Templating_0* and *SQLQuery_0* being connected into the process flow. A component that generates a word cloud is being connected to text box display component (*TextBox_0*) and a search capability (*PubMedSearch_0*). The graphical diagram on the right of the screen shows the components (including users and developers)

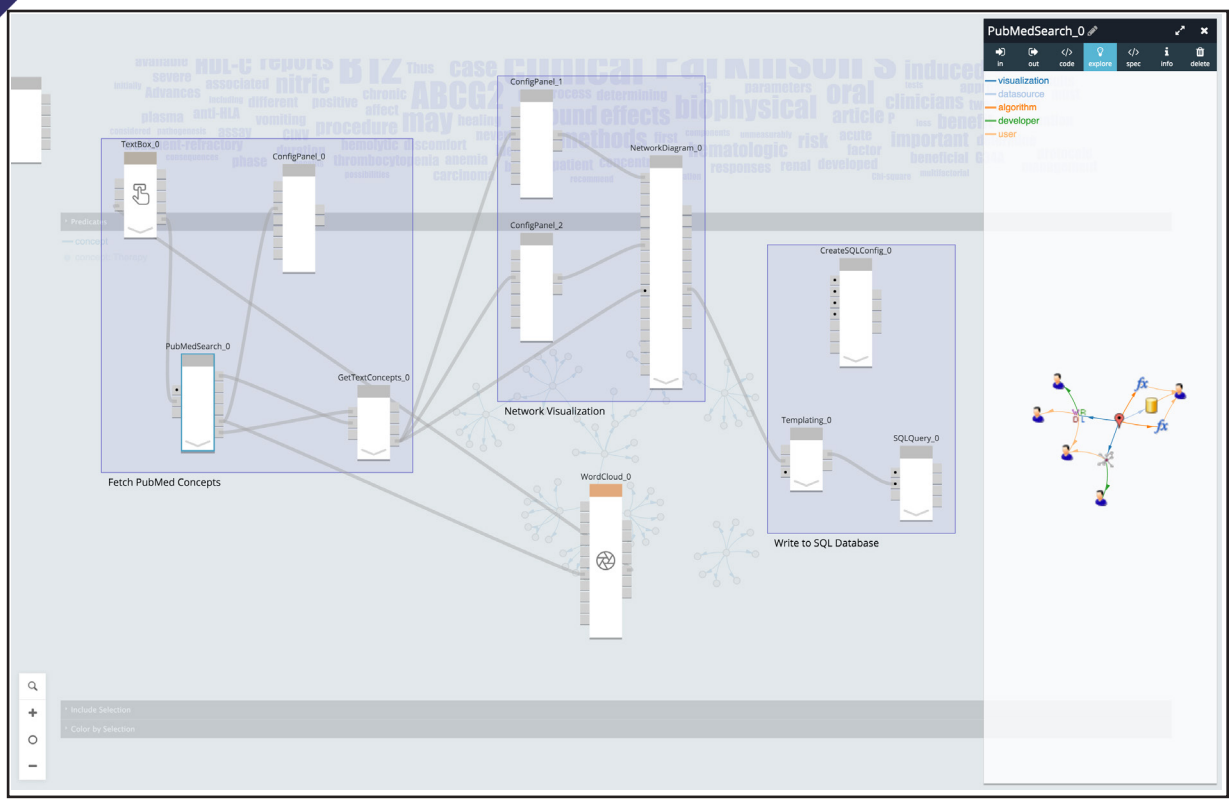

Figure 1. The Drag and Drop Interface

and their connections.

An important point to understand, and one of the most compelling features of the Exaptive development environment, is that the stubs on both sides of the objects refer to methods (i.e., functions that can be invoked). Simply click on the stub, and it will display its function. When you graphically connect a stub that creates output from an object to a stub on another object, it accepts it as input you are programming into the data flow.

With Exaptive, you can designate any executable program or application to be an object. It provides an API to other commonly used IDEs, such as Eclipse, which may be needed for coding new software objects from scratch. Of course, the software comes with a standard set of components and component templates.

Navigating Through the Data Landscape

Providing an environment where the developer and user can navigate their way through large amounts of data requires technology that goes beyond traditional metadata directories. Exaptive employs the Entity, Attribute, Value (EAV) model for data representation, as this is well suited for statistical data. It also makes use of a user-driven ontology (a semantic reference structure) to define data sources. Taken together, this means that data relationships and meanings are visible to the user and can be effectively employed to navigate, explore and process multitudes of data sources.

The ability to define, specialize, develop and deploy software objects means the developer, particularly the data scientist developer, is provided with a full palette of analytical capability, allowing for the exploration, gathering, cleansing and analyzing of many data sources.

Collaboration: The Xap Store

If Exaptive customers were working in isolation, they would develop their own specialized sets of components to add to the initial set the product provides and move forward on that basis. However, Exaptive quickly realized that its rapid development platform would prove far more useful and be able to address a much broader set of applications if there were a healthy population of prebuilt and proven components. To make this possible, they created the Xap Store – a marketplace for data applications.

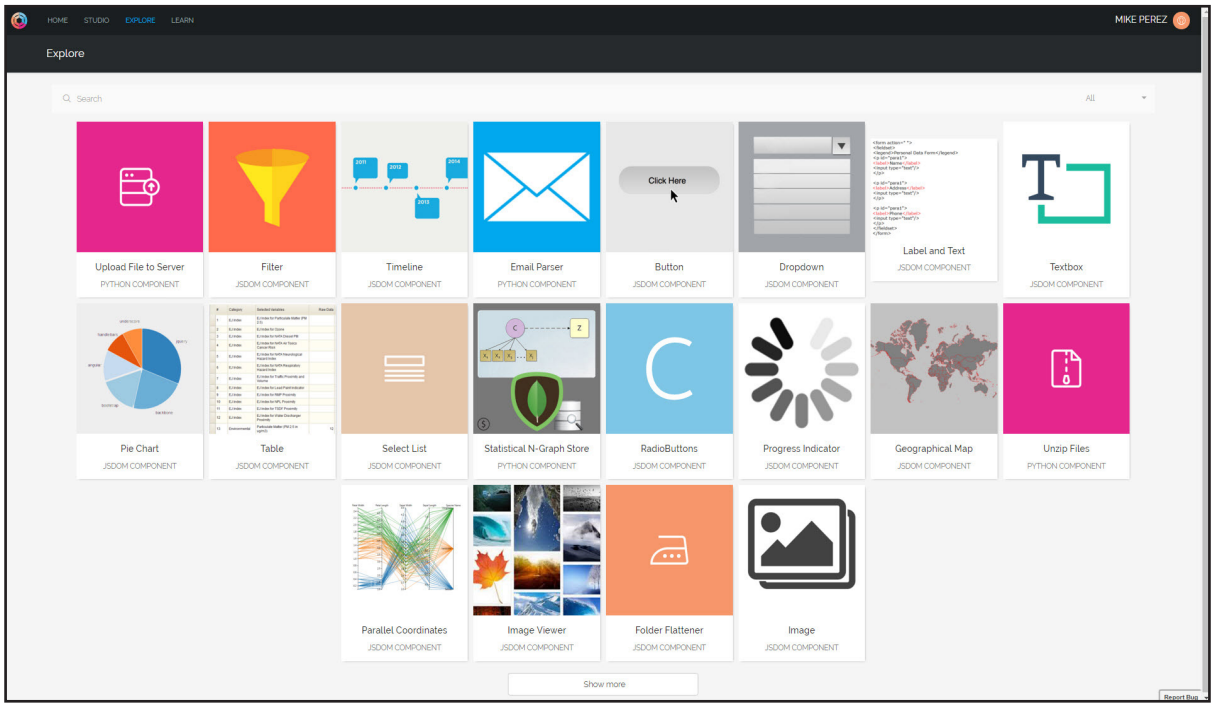

Figure 2. The Xap Store

The Xap Store, as shown in Figure 2, is not that different from the app stores of Apple of Google, in the sense that some of the apps are available for free and some have a price attached. To use the store, you register and download the Xap, after paying the appropriate fee, if there is a fee involved. The user can obtain modular components (i.e., ready-made Xap objects), collections of related components or full blown applications that are ready to run, including in some instances data or links to public data sources. Users can also create and post their own components or applications.

There are many aspects of the Exaptive platform that are impressive: its visual interface, its development style, its array of data interfaces and its approach to data navigation. However, in our view the collaborative environment that the Xap Store encourages is the jewel in the crown. It enables both software engineers and data scientists to share or sell their creations for the benefit of the whole community. The result is that Exaptive's capability naturally expands with time.

The Xap Store is more than just a software marketplace and collaboration zone; it's also a destination for developers to visit when they wish to build solutions for specific problems. They may be able to investigate, for example, which particular machine learning algorithms have been applied to their specific focus area or how a particular analytical solution has been implemented on Hadoop or Spark. As such, it is not just about sharing compatible software components, it is also about sharing ideas and pushing the boundaries of what is possible.

The Right Tools, the Right Staff, the Right Business Processes

Since the early days of the Big Data trend, we have witnessed an incremental increase in the sophistication of technology. It has been clear from the start that no single software product or environment would be able to satisfy the burgeoning and increasingly diverse requirement for analytics. Part of this has to do with history. Organizations that already deployed a sophisticated analytics capability did so on older data warehouse technology using traditional analytics tools. They were not going to throw away the capabilities they already had, but neither would they ignore the new capabilities: scalable parallel Hadoop clusters, Spark, open source machine learning libraries, the wealth of capabilities available in R and Python libraries and so on.

Companies new to analytics were inevitably going to adopt the newer, more scalable and less expensive capabilities, especially when they could do so using the cloud. And of course, they did. From a software tools perspective, the only capability that could span the complete set of such a diverse landscape of technologies would be a platform that worked in a service-oriented manner, able to thread together the technical capabilities that would be needed to connect all data sources and, where necessary, store the data in scalable and affordable data stores. This is what Exaptive chose to provide.

In regards to the skills situation, which also will vary from business to business, there is nevertheless a need to provide a development platform that accommodates the efforts of both technical level developers and data scientists. Such a platform is most effective when it can be used in a highly productive iterative fashion to build both batch and real-time analytics capabilities as well as implement those solutions into operational systems. This, too, is something that the Exaptive platform accommodates. Clearly, it is not a low-level programming environment, but it enables low-level components (objects or services) to be clipped into the execution environment. As such, it is language agnostic. Components can be built in C++, Java, Python, R or whatever – the appropriate language and program development environment can be clipped in as necessary.

This is about building data applications rather than dashboards. The process for building such applications is naturally iterative and involves discovery, experimentation, iteration and implementation – including, where necessary, data transformations, data cleansing, data enrichment and so on. Such activity is, in our view, a natural aspect of the essential R&D process of an organization.

Since the turn of the century, a long procession of web businesses, social networks, gambling sites and many multi-player gaming platforms have established themselves. What these businesses have in common is that they are entirely data driven, and their R&D activities consist almost entirely of data analysis. Other businesses in other sectors have taken note. It may not have previously been economic to explore and analyze all the data of an organization, but this is now increasingly becoming possible – and such data is capable of providing a complete map of all the business processes of an organization. The exploration and analysis of such data constitutes research into the business and will almost certainly provide actionable knowledge that can, with the right software platform, be put to use instantly.

The Power of Cognitive and Predictive Analytics

Business applications fall into two distinct categories: those that automate the activities of the organization (the OLTP and office applications) and those that provide feedback on business processes of the organization (the BI and analytics applications). Naturally, business automation applications evolved first, with feedback applications following in their wake. Early computer

systems had fairly primitive BI capabilities, consisting of relatively inflexible reports. These gave way to dashboards, OLAP drill-down capability, performance management and rich data visualization. BI applications reached their zenith with the establishment of data warehouses and distributed data marts, enabling tools like Tableau and Qlik to provide an individual BI data service to users.

Analytics, previously a niche activity, recently burst out of its nook with the advent of Hadoop and its extensive open source ecosystem, enabling very fast and inexpensive analytics applications to access large data volumes. Analytics is also best viewed as having two categories: predictive analytics that report on real-time trends and cognitive analytics that explore data to discover new actionable business knowledge. With predictive analytics, the technical challenge is to process streams of data whereas with cognitive analytics, the challenge is to process large volumes of data.

The term "cognitive analytics" is relatively new and is worth examining. It is generally used to describe the use of Big Data gathered from diverse sources to generate business insights about specific business areas such as customer preferences or marketing effectiveness. Much of the complexity in this activity resides in preparing the data: gathering it from multiple sources, capturing metadata, cleansing it of spurious data and transforming it into a structure that is suitable for the analytical algorithms that will process it. The analytics produces results which are then pondered over and will likely lead to further iterations of the process.

The title data scientist was born from this. The activity is similar to the scientific process of formulating a hypothesis, analyzing the data and then forming a tentative conclusion in light of the results. This will usually lead to adjusting the hypothesis, perhaps gathering more data and applying more analysis. The process repeats until the evolving hypothesis is confirmed. The result is knowledge, which can be used effectively by the business.

The knowledge discovered may simply be a rule that can be applied to improve a specific activity, the discovery of specific thresholds when certain outcomes prove likely or the identification of a particular trend that is worth monitoring. If we take, for example, the often-discussed situation of customer churn in the telecommunications industry, it may be discovered that customers in a particular geographical area rarely change providers because the competition has poor coverage. This rule will help determining pricing in that area, and the customer churn rate in the area needs to be monitored to determine whether the rate changes if new pricing is introduced. Alternatively, it may be discovered that a certain number of dropped calls within a given time span provokes customer churn, so this establishes a threshold that triggers when service levels need to be improved. Or it may be discovered that trends in churn are strongly affected by competitive marketing activity, and thus predictive analytics needs to be employed to identify those situations and respond to them by, for example, tactical competitive service offerings.

An interesting aspect of this is that cognitive analytics will often lead to the development of predictive analytics capabilities, which are likely to become an integral part of operational systems.

The Nervous System in Man, Analytics in the Business

We can see analytics as similar to the way that human beings learn sophisticated physical tasks, such as driving a car. While you may remember this as being a relatively swift process, it is complex, and in the early stages, it is slow.

First you learn the controls (you gather data) and you learn driving activities (using the accelerator, the brakes, steering, etc.). You normally execute this in a relatively controlled environment. Initially, all of this is cognitive in the sense that the brain is making all the decisions. You learn by integrating data from different sources: your own nervous system, the behavior of the controls, the readings of the instruments, the movement of the car and so on. At a certain point, the speed at which you can incorporate what you have learned increases significantly. The cognitive control of driving activity that the brain has been providing is operationalized by the body.

But in such a complex activity there is a great deal to learn, and the learning process (the cognitive function) continues to operate and gather data. As this proceeds, what is being learned is gradually converted into operational behavior. The analogy here between human behavior and analytical systems is almost exact. What is learned about the behavior of the car – accelerating, braking, steering, etc. – is implemented predictively by the autonomic nervous system. You learn with increasing accuracy how to predict the behavior of the car, the behavior of other road users, the variety of different driving conditions and different types of road. Cognitive analytics is implemented as predictive analytics, just as it is in the business.

And just as predictive analytics applications are not static, neither is one's ability to drive a car. Indeed your inner "driving system" will need to be adjusted when you drive a different car, when you drive a car in a foreign country or if, for the first time, you drive in icy or snowy conditions.

In time, the analytical systems that a business builds become an integrated part of a business behavior, and the systems are unlikely to remain static, unless the internal systems of the business sector it occupies become static.

Exaptive in a Business Context

An interesting and very prominent area for the application of analytics is in the healthcare environment. There are many reasons for this aside from the impact of legislative changes. The fact is that more and more data is being captured or created and then shared by health organizations. This has the potential to generate new knowledge about treating specific ailments effectively and economically. However, the abundance of data alone does not guarantee useful insights

Exaptive has been put to use for this purpose, and how it has been used merits discussion, not only in respect to the data it has been applied to but in how the analytics process often proceeds. Even if the data has already been conveniently gathered, cleaned and transformed for use, the data-driven discovery of insights is a not as simple as selecting a few algorithms and marching them through the data. This is only likely to be the case in frequently analyzed data environments and business processes or when re-examining precious work. More often, it is a process of discovery, as it was in the cancer research use case that we now discuss and illustrate.

Exaptive Use Case: Leukemia and Gene Expression

A cancer researcher was examining a collection of patient data in the hope of discovering patterns that might identify causes or co-factors in the onset of leukemia. Figure 3 depicts a screen shot of the completed project with all its graphical elements. The process that created it involved examining and linking together three data sets from different sources and in different formats (some proprietary, some public). The researcher has processed them using a statistical package, then visualized results in a series of interactive visualizations as possible paths of

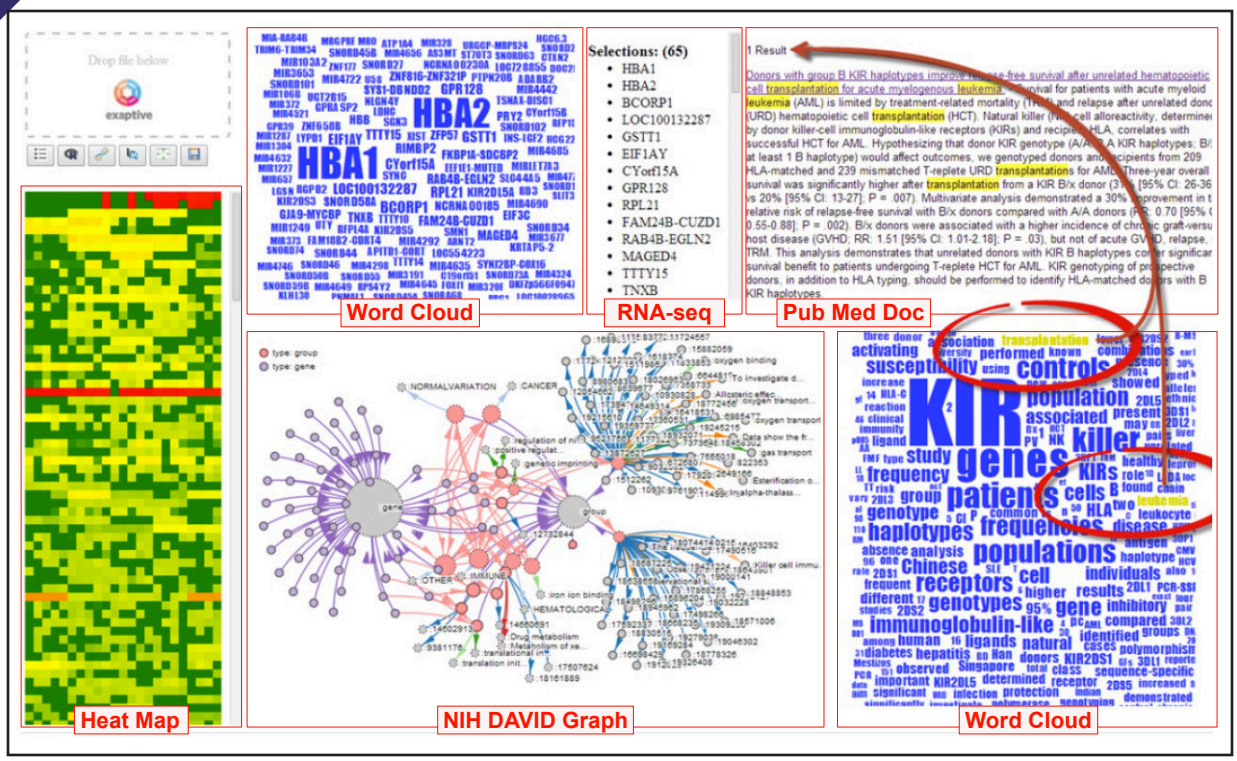

Figure 3. Cancer Research Use Case

investigation presented themselves. The result shown only illustrates the choices that were taken, ignoring blind alleys that might have been pursued.

The result was arrived at in the following way. At first, the researcher employed some programs written in the R language to analyze the RNA sequence from patient data, which included gene expressions for 23,000 genes from 23 different patients (labeled RNA-seq in the diagram). An interactive heat map was generated (labeled Heat Map in the diagram) to highlight the frequency of occurrences of specific gene sequences. This was then linked to a word cloud (Word Cloud in the diagram).

These processes provided the researcher with an immediate picture of the data, allowing him to home in on 65 genes with the greatest variation in expression among the data. At this point, the researcher chose to query the NIH DAVID database for annotation information about clusters within the 65 genes, creating a further data visualization (the NIH DAVID Graph). This graphical Gene ID mapping allowed the researcher to discover a cluster of 5 genes that exhibited a cancer annotation and also indicated that the difference between the cancer variation and the normal variation was caused by a single gene in the cluster.

Next, the researcher was able to see from PubMed annotations contained in the NIH DAVID database that two clusters had significant published research associated with them. So, after narrowing to certain gene clusters, the researcher examined the journal literature on one of those clusters available from PubMed. This led to the creation of a second word cloud, which displayed the occurrence of terms in the PubMed document database. The researcher was interested in the documents that mentioned leukemia and transplantation, and so by clicking on those terms in the word cloud, he was able to filter through the list of articles until the single specific document shown was found.

What this use case demonstrates is the dynamic interactive nature of the Exaptive platform. It enables the user to access the wide variety of tools that can be put into the Exaptive toolbox.

The tools can range from full blown implementations of machine learning algorithms to the usual practical visualizations, such as a graph, heat map or word cloud, as illustrated in this example. The platform is built for interactive iterative analytics, and, in our view, data science is most effective when conducted in this fashion.

Exaptive has already been employed extensively in healthcare and the life sciences:

- Enabling the collaboration between a consortium of organizations to research brain disease (primarily solving a data integration problem)
- Swiftly identifying potential false negatives in multiple sclerosis diagnosis, reducing the time taken from months to days (primarily solving a latency problem)
- Enabling a biobank to navigate through dozens of disparate research studies using its biosamples, demonstrating it to be perhaps the richest multiple sclerosis data set in the world (primarily solving a data integration problem)
- Empowering geneticists to identify new connections between genes and tumors to design breakthrough personalized cancer diagnostics (primarily solving a data volume problem)

Business Summary

One of the problems that has bedevilled many of the Big Data analytics initiatives is the friction that can arise between software engineering staff, whose priorities lie in building reliable software infrastructure to deliver a data service, and BI and analytics users, whose priorities are to pursue projects and implement capabilities quickly.

Exaptive can diminish this friction for three reasons. First, it is not solely an analytics tool. It is an object-oriented/service-oriented platform capable of use in many contexts, including software engineering contexts. Second, it is a fast development environment no matter who is using it. Third, it is fundamentally collaborative.

The point is that there is no reason why both sides of the data science/technology divide cannot use it collaboratively. From the data perspective, it is technology agnostic in the sense that it can provide access to data from remote sources or local sources: databases, data lakes, data warehouses or file systems. Moreover, it is engineered to provide a graph-based connectivity architecture that enables and encourages the ad hoc inclusion of additional data sets from any source adaptively, regardless of schema. It is focused on interoperability.

From the software perspective, it is inclusive, supporting the customary tools of the data analyst and the software engineer (the Hadoop environment, SQL, SPARQL, R, Python, Javascript, etc.) plus all the capabilities available through the Xap Store. And of course, it is browser-based and hence very easy to deploy in multiple contexts. The Exaptive platform thus accommodates both data scientists and software engineers, facilitating collaboration between the two disciplines at an individual and team level.

In our view, analytics will become a fundamental component for the vast majority of businesses. It embodies the implementation or evolution of cognitive and predictive analytics within the business to the point where they become broadly applicable embedded business processes. For this and other reasons, Exaptive should be attractive to organizations that realize the importance of this technology.

At the interface level, it is purpose-built for data science activity and is consequently easy

to use. It facilitates data integration, providing for genuine data self-service. It enables the individual data scientist and the company to build and implement analytical applications and gradually expand both skills and analytics tools. It accommodates all staff involved in the data science activity: the data scientist, the business analyst, the line manager, the end user and the software developer. It provides a marketplace that can be explored for new capabilities and new application ideas.

Given that data scientist skills are in short supply and expensive, the fact that Exaptive enables rapid development, significantly shortens the discovery process and improves data scientist productivity leads to faster time to value, which may be its most important feature.

About The Bloor Group

The Bloor Group is a consulting, research and technology analysis firm that focuses on open research and the use of modern media to gather knowledge and disseminate it to IT users. Visit both www.TheBloorGroup.com and www.InsideAnalysis.com for more information.

Austin, TX 78720 | 512-524–3689