Innovation Winners Create “Situational Awareness” with Big Data Analytics Platform

Innovation hurts. It’s messy. And if not managed “right,” it can be very expensive. That’s why CIMS believes the winning firms will be those that put a premium on acquiring timely and relevant market/business intelligence to guide their 24/7 decision making. It’s what we call “situational awareness.” Coupled with the power of predictive analytics and the new Big Data Analytics Platform I describe here, it will put knowledge of impending market shifts, emerging competitors, as well as potential partners, into the hands of every business-function manager. This, in turn, will empower them to make the “right” business decisions. Here’s how:

CIMS believes that innovation is best led by market-savvy business-function managers working through aligned and empowered cross-functional teams. However, no matter how talented and experienced these managers are, they face a daunting challenge because of the speed and volume of data being created every day on the planet in the form of web pages, blogs, wikis, and so forth.

We estimate that over 15 petabytes ($10^{15}$) of data are being added to the worldwide web daily. Furthermore, 80% of this new data growth is unstructured content, i.e. email, blogs, web pages, white papers, images, video and audio. In fact, the Internet is now estimated to contain over 1.2 zettabytes ($10^{21}$) of data. That’s roughly 60 million Libraries of Congress!

The challenge, therefore, becomes separating appropriate business signals from the petabytes of “noise” generated, both inside and outside the organization.

Figure 1 – Big data requires big storage. Expect 20-plus terabytes ($10^9$) of storage today to grow exponentially in the future.
Situational Awareness Is Key
For the past two years, CIMS researchers have worked to demonstrate that situational awareness on this scale, is not only feasible, but when embedded in a culture of innovation, it can be the source of competitive advantage. Advances made at NC State University in cloud computing, along with advanced text analytics tools supplied by IBM, now make widespread implementation of this type of business decision-making feasible. To prove this new method of business intelligence is generic and can be applied successfully in any business or R&D activity, we and our partners (IBM Software Group and the Department of Computer Science) are developing and testing an ‘industry-standard’ Big Data Analytics Platform (BDAP™)1 across an array of strategic business issues.

BDAP defined
The BDAP platform embraces four major components:

I. Cloud computing environment
II. IBM Analytics Software
III. Business Process (gathering, filtering, annotating, and displaying massive amounts of data)
IV. Resulting Meta Data File (MDF)

The cloud computing environment is hosted by the Virtual Computing Lab (VCL) at NC State University. NC State’s Department of Computer Science has been a pioneer in cloud computing since 2004 (see Chasing Smarter Clouds: CIMS Technology Management Report Winter 2010 – 2011, p.1 http://cims.ncsu.edu/index.php/newsletters). In the VCL, we use two cloud configurations, one that serves our students and faculty, the other a sub-cloud called the High Performance Computing Center (HPCC) where we perform our crawls and annotate the data (see figure 1). In the HPCC we have specialized processors with lots of local memory for each processor. This is essential for “mashing up” and annotating millions of records at any one time.

As you can see in Figure 1 big data analytics requires lots of storage - 20 plus terabytes today and, with the help of the major storage equipment suppliers, exponentially more in the future.

Analytics Software
IBM provides the software platform shown in Figure 2. It is based on natural language processing of free form text and is the same set of programs that delighted Jeopardy fans when Watson easily beat the two all-time champions! Of course, we are not interested in the same categories of information that challenges TV audiences. We scan the web looking for things like the formation of new ventures, the completion (or failure) of a clinical trial, new laws and regulations, federal grant solicitations, even consumer sentiment regarding a

1 Trademark is pending the completion and approval of NCSU application by USPTO.
product or a service gathered from Facebook and LinkedIn. From this we can piece together a powerful mosaic for a company. For instance, we have been able to identify a potential business partner for a leading biotechnology organization by analyzing over 400 websites, 9.5 million web pages, and 4.5 terabytes of data.

Our Business Process
This brings me to the third, and perhaps most critical component of the BDAP, the business process to create and use big data. Next to each of the software elements in Figure 2 are the major activities associated with each step in the process. Also shown are the resources required by each activity.

The BDAP process is CIMS’ main contribution. Moving business to think beyond the data associated with internal enterprise resource planning (ERP) software tools and structured database analysis requires a change in the way one mines data. Understanding how to ask business questions properly is of paramount importance when dealing with big data.

This will require individuals to look beyond standard statistical approaches and begin to explore other aspects of stakeholder information such as social networks, unstructured data sources, as well as internal data sources, knowing that previous limitations based on data type can now be overcome. In fact, all of these data need to be “mashed up” to yield the fourth component, a larger, and potentially proprietary, Meta Data File that will drive decision-making.

Where we are headed
The BDAP is not done; in fact we are just starting. At the CIMS Corporate Sponsors Meeting, this past May, IBM announced that they would add their suite of SPSS products onto NC State’s VCL which enables us to do predictive analytics. Having a Q & A interface may be fine for Watson, but not for businesses. They want to be in front of their competition and know when market conditions will likely change. For example, product developers are keenly interested in information that signals when populations of consumers are moving from “early adoption” to more fully “adapting” a new technology. Identifying this critical
inflection point, commonly referred to as “crossing the chasm,” can make all the difference in a company’s profitability and brand reputation.

**Movies Too!**
Closely linked to predictive analytics will be our pursuit of new visualization techniques. Indeed, we are finding that they are inseparable. Analytics helps create more effective visualization, and visualization assists analytics by allowing users to “see” the data. For example, by crawling the same websites periodically, say monthly, and only keeping the data that have changed - we can create “movies” of certain information over time. These patterns are vital for discerning trends and would be impossible using other analytical methods.

**Learning BDAP**
In parallel, we are developing BDAP training courses for our MBA and MS students as well as short courses for CIMS member companies. Smart, energetic students are the mainstay of NC State University. We believe in exposing them to the latest management techniques before they go out into industry. CIMS views this as a takeoff of the successful strategy Apple has used to capture the hearts and minds of K-12 students. Graduate students will learn to apply BDAP in companies as part of their practicum projects. And when they go out into business they will demand this new business intelligence capability that they helped to develop.

**What you can do now**
Similarly, business-function executives must learn how to harness this new capability. Using an action-learning format, they will learn how to apply BDAP to their own organizations. Working with IBM, we have developed a five-step engagement model that allows companies to progressively build proficiency and confidence in big data analytics as they learn to apply it to decisions in their organizations (see “Engagement Model” in *Business Intelligence “in the Clouds”*, on p. 5, TMR Winter 2010 – 2011 [http://cims.ncsu.edu/index.php/newsletters](http://cims.ncsu.edu/index.php/newsletters)).

We have also developed a set of workshops, in both open and custom configurations, that shows business function executives how to use big data to innovate new business models, products and services.
**Custom Solutions for You**

If your organization is typical, it is already being bombarded by consultants, data aggregators, and commercial software tool vendors who claim to do the same thing. Don’t be fooled by these comparisons, which are shown in Figure 3. CIMS is not selling a tool--we are building an entire platform that allows you to build a Meta Data File that is customized to your business needs. It is one where: 1) you ensure the quality of data by selecting just those URLs you have determined to be reliable and credible, 2) you meld the web-based information with your firm’s internal operational data files, and 3) you annotate and display the data in a way that supports your specific business processes and analytical needs.

BDAP will provide clarity to an organization’s decision making. As importantly, it will create a new ‘situational awareness’ that will align your workforce and accelerate innovation.

If you believe your organization would benefit from having such an information advantage, or if you just have questions about any of the topics addressed in this article, please don’t hesitate to email me at Paul_Mugge@ncsu.edu.