HOW YOUR BUSINESS WORKS
information management
July 19, 2012

Breaking the Barrier of Big Data Analytics in BI

By Todd Nash

Before the era of computers, databases and, now, big data analytics, business intelligence was based on one's gut instinct and the creative vision necessary to capitalize on it.

A June 2012 Chicago Tribune article by Phil Rosenthal entitled"This Man Knew How to Size Up Consumers" highlights the 40+ year career of David Wallerstein to illustrate this point. Mr. Wallerstein, a friend of Mr. Rosenthal's grandfather, was notorious for telling outrageous stories involving his influence with Walt Disney and Ray Kroc, among other notable entrepreneurs. After his death 20 years ago, friends discovered that he didn't make it all up. His obituary backed up his stories; for instance, as a long-time operator of movie theaters, he was the man who introduced butter on popcorn, as well as ice in soft drinks. As a board member of McDonalds, he also convinced his friend Ray Kroc to add large size fries to McDonalds' menu. Mr. Wallerstein was seen as a brilliant pitchman who relied on assimilating his creative instincts with his observations of consumers' tastes and behaviors.

Fast forward to today — our global business environment is too large to depend solely on intuition. Now business intelligence, which can tell us what consumers are buying and what influences them, is obtained at warp speed through the analysis of mass quantities of both structured and unstructured data — big data. What's the new constraint? Creativity, which Mr. Wallerstein had in spades. In today's data-rich society, business users grapple with how to fulfill the promise of big data through innovative approaches to analyze and leverage it effectively.

As far as big data is concerned, businesses could use more of Mr. Wallerstein's creativity, particularly relative to big data. For years business users have been saying, "The things I could do if I got my hands on all of that data!" But throughout this period, too many technological and financial barriers stood in the way of big data's potential. From acquiring too much data to store and process to the fact that the data was inherently unstructured, the algorithmic gymnastics required to integrate it were extremely difficult and cost-prohibitive. With all of the recent advances in technology, businesses can now remove these constraints and take advantage of big data in a new way.

What is Big Data?

In today's global marketplace, deriving higher quality solutions to solve challenges in business, health care, finance, education, agriculture, retail and other industries now requires the power of big data analytics. The promise of big data analytics is that it can organize, quantify, and make sense of virtually all types of random data deviations to help decision-makers better understand an enterprise's needs and goals.

Big data was so named because of the sheer volume of data currently being processed throughout the computing universe. While many disparate types of data have been around for years, until now this raw information hasn't been harnessed or structured effectively. As our society evolves toward the use of more technology (PC, tablets, smartphones), more data is generated (e-shopping, e-transactions, e-reviews, social media, etc.) that could be used to great benefit. In the past, limited technology, combined with high costs, detracted from an enterprise's ability to process data and design products to analyze it; it had to be structured within an inflexible database format and the costs of programming or developing a new analytic structure were so high that the expense was rarely justifiable or profitable.

Why is big data so prevalent now? The "Aha!" moment occurred as the market found a way to leverage commodity hardware and software technology, helping enterprises process this information quickly, cheaply and effectively. For example, now a terabyte of data can be stored on a flash drive. Processed data can be reduced from thousands of dollars per byte to hundreds, or even tens of dollars. Data mashups can

be executed on the fly in reporting tools. All these advances allow enterprises of all sizes access to big data analytics so that they can make more focused, informed decisions.

Big data can be classified into one of three types: structured data, semi-structured data or unstructured data.

See Table 1, Right

Creative Tools and Techniques to Gather, Process and Integrate Big Data

It's important to remember that it took David Wallerstein many years of first-hand observation to understand his consumer target market before acting on that knowledge. Like Wallerstein, most human beings have a similar cognitive capacity to store, retrieve and process seemingly disparate bits of unstructured information and events, form conclusions from this information and then make educated decisions. The trick in today's big data world is to use that same creativity to develop rule-based toolsets that can rapidly do what our brains do naturally: store, retrieve, process and integrate all of the structured, semi-structured and unstructured information that is currently available, to form an analysis and make educated decisions which can further an enterprise's goals.

Each of the data classifications defined in Table 1 requires specific toolsets to process, organize and combine it into a fully integrated big data database and drive out valuable analytics.

See Table 2, Right

Imagine the possibilities if someone like David Wallerstein had these dynamic big data analytic tools at their disposal. For example, one innovative toolset for semi-structured and unstructured data is comprised of a rules-based engine which converts voice to text from video and audio data, and puts it into semantic context for integration with more traditional, structured data. An enterprise can capture completely unstructured content, run it through bundled commodity PC hardware at lightning speed, and create a data-rich treasure trove that is fully compatible for integration with valuable business applications.

It's also interesting how what was considered old is now new again. If you evaluate Hadoop and its associated capabilities, it is very similar to the old VSAM file structures and Pascal programming language. Hadoop provides tremendous flexibility and processing power, but requires strong programming capability and deep thought in how to process this unstructured data. The beauty of Hadoop and other similar modules is that they run on any platform, using vast resources of bundled, standard commodity PCs, cost-effectively taking advantage of a tremendous amount of

Data Type	Structured Data	Semi-Structured Data	Unstructured Data
Definition	Traditional data in a traditional database structure	Semi-structured data has some format to it, but also has unstructured "blobs" or "gold nuggets." This information can't be processed by standardized tools and requires more sophisticated tools to provide context and rules to process it effectively.	Unstructured data is not consistent from one record to the next. Defining a record is even challenging. Most of the content on the Web is unstructured; the nature of this data requires new technology to process, count, evaluate or filter it so that organizations to take advantage of this data.
Examples	Enterprise resource planning; transaction processing systems; customer relationship management processes; backup storage for large data volumes; data repository	Spreadsheets, including budgets, baselines, commissions, data stores; word processing documents, including contracts, performance evaluations, legislation, or surveys; call center logs including toll-free responses; and Web logs that track website activity	Customer reviews, chats, blogs; Facebook or LinkedIn relationships; visitor behaviors; maps including locations and geo-spatial information; weather data, weather maps; surveillance audio/video, advertisements, testimonials, YouTube videos

Table 1

combined processing power. Hadoop is installed on a node or cluster and can process tons of data, using simple code to filter out superfluous bits that don't apply and to store bits of data that do.

Data Type	Structured Data	Semi-Structured Data	Unstructured Data
Toolsets Used:	***Traditional databases*** (Oracle, DB2, SQL Server), Integrated Data Warehouse Appliances (Teradata, Netezza, Exadata, MSFT Parallel Data Warehouse) and software Data Warehouse Appliances (Vertica, Greenplum, Kognitio, SAP HANA) that can leverage commodity hardware. These databases are built specifically for high performance capability, high volume, and high workload management capability. Many have integration with Hadoop to take advantage of the pre-processed, unstructured data.	***Textual Extract, Transform, Load*** *tools* (Textual ETL, Forest Rim Technologies, traditional ETL tools with bolt-on rules based engines). They build context and semantics layers along with rules based engines to pull out the richness of that data "nugget," measure it and put structure around this unstructured content. Once it's measurable, it can be integrated with traditional structured data to provide much richer analytics.	***Hadoop (or an equivalent file-based storage structure)*** is used to capture this unstructured, unruly data as defined in *Table 1*. Hadoop combines more traditional data storage structures with data processing toolsets, similar to older programming languages, such as Pascal or Assembler.

Table 2

However, processing the data is only the first part of big data's creative challenge. The next piece of the puzzle is to create an effective reporting tool that has more than just the ability to build reports. This is where another creative leap in big data comes in, involving data visualization toolsets (such as Tableau, Spotfire, QlikTech and many others). These tools are characteristically simplistic in nature and easy for a typical business analyst to pick up quickly. Data visualization toolsets provide a simple way to build dashboards and graphics and, most important, provide a way to combine data from a data warehouse with other data that may be department-specific. All the tools described here are here

to stay and have significantly improved businesses' ability to interact with data without being dependent on IT support.

How Can an Enterprise Take Advantage of Big Data?

In the past, business users who weren't blessed with David Wallerstein's insights or friendship were limited to executing analyses on limited types of transactions, typically revenue, to understand customer behavior, trends and key performance indicators. Enter today's global, highly competitive marketplace, which relies increasingly on quality BI to make decisions that will ensure its survival and growth. As technology constraints are eliminated and the associated cost of technology is reduced significantly, other types of data can be leveraged:

- Transactions: non-revenue based data, such as reward points, warranty policies, etc.
- Interactions: call center data for customer touchpoints; website/portal requests to understand the products or services that customers/prospects are interested in; survey responses for feedback
- Research/Investigation: website interactions
- Sentiment: reviews, blogs, social media, etc. to understand positive, negative and neutral sentiment.

All of these big data advancements provide a window of opportunity for enterprises of all sizes, in all industries, to take advantage of this available technology. The next step is to take this treasure trove of now-structured, integrated data and place it into proven business applications. These business applications create a more comprehensive visual picture of opportunities to make decisions and, consequently, derive real business value from these choices. This is precisely where business needs to get creative in determining how to leverage this newfound capability and company asset.

Several examples of the use cases that are driving business value include:

- Brand/Product Management — to track consumer sentiments and integrate them with the transactions in order to understand positive and negative drivers.
- Product Flexibility — to quickly respond to market trends by refining products or services that more accurately reflect the needs, demands and transaction patterns of your customers or potential customers.
- Marketing Lift — by identifying your positive and negative targets and campaigning directly to their needs or demands.
- Risks and Fraud Reduction — by detecting fraudulent transactions and determining the characteristics of the transaction, establishing the culprit and any pre-work the culprit was doing to set up the transaction.

David Wallerstein may no longer be with us, but big data analytics have come a long way in providing the creative business intelligence his genius so intuitively understood. Technology isn't the limiting factor it once was and costs are much more reasonable, so the use cases are growing by the day. These ideas are only limited by the business user's imagination. It's time to get creative and answer the call of "The things I could do if I got my hands on all of that data!"

Todd Nash is a founding Principal at CBIG Consulting, a professional services firm that helps clients leverage their data assets through big data analytics to produce timely, effective business strategies and tactical decisions. Nash's responsibilities include managing, developing and implementing big data analytics architectures, business intelligence/data warehousing solutions, and cloud-based analytics services. CBIG's subject matter expertise extends to a variety of industries including consumer packaged goods, telecommunications, manufacturing, insurance, banking and health care. You can reach Todd at todd.nash@cbigconsulting.com.