

Getting Big Value From Big Data... Fast

An ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) White Paper
Prepared for Tableau Software

May 2011



*IT & DATA MANAGEMENT RESEARCH,
INDUSTRY ANALYSIS & CONSULTING*

Getting Big Value From Big Data... Fast

Table of Contents

Executive Summary	1
Understanding Big Data.....	1
Managing Big Data.....	2
Getting Big Value from Big Data.....	3
Tableau Software Technology.....	4
EMA Perspective.....	5
Customer Perspective	5
Final Thoughts on Big Data.....	6
About Tableau Software	6

Getting Big Value From Big Data... Fast

Executive Summary

The exponential growth of data is putting a strain on traditional Business Intelligence (BI) and analytic solutions. New strategies to manage this information and leverage greater business value are top of mind with both business users and IT management. Supporting the needs of self-service users and connecting to a wider assortment of data adds hurdles to achieving a stronger BI environment, but are critical for success. This ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) report defines Big Data, identifies the necessary components to better manage it, and helps the user understand how to create BI value through Big Data.

Understanding Big Data

Over the past decade data growth in the enterprise has expanded at an exponential rate. The data that companies manage has quickly evolved from gigabytes to terabytes to petabytes with no limits in sight. In 2005, WinterCorp, a leading expert in large database management issues, identified the world's first 100-terabyte data warehouse.¹ It came as no surprise that an Internet company, Yahoo had set this new standard. Only six years later Wal-Mart, the world's largest retailer, is logging one million customer transactions per day feeding information into databases estimated at 2.5 petabytes in size.² The Large Hadron Collider at CERN the European Organization for Nuclear Research can generate 40 terabytes every second during experiments³ and Boeing jet engines produce ten terabytes of information every 30 minutes of operation.⁴ A four-engine jumbo jet can create 640 terabytes of data on just one Atlantic crossing. We have officially entered the Big Data era of computing.

The term Big Data has yet to be universally defined, but generally speaking Big Data represents data sets that can no longer be easily managed with traditional or common data management tools and methods. The velocity and scale of Big Data brings new challenges to access, search, integration, discovery and exploration, reporting and system maintenance. This new age brings with it new data sources that are adding stress to existing infrastructure. This is especially true of real-time data and unstructured data. Sensor and machine data are leading the way in overall data growth as highlighted in the Boeing jet example above. Internet commerce sites generate massive data tracking consumer behavior in ways that was not possible in an earlier age. Social networking data is a new and highly valuable source of information for the enterprise, but again its sheer volume and speed make it difficult to utilize. The Micro-blogging site Twitter serves over 150 million users who produce 90 million "tweets" per day. That's 800 "tweets" per second. Each of these "tweets" is approximately 200 bytes in size. On an average day this traffic equals 12 gigabytes and throughout the Twitter ecosystem the company produces a total of eight terabytes of data per day. In comparison, the NYSE produces just one terabyte per day.⁵ Social data brings with it new data types such as location, behavioral, sentiment, social graph and rich media data that all represent new value to the enterprise. Big Data is fueling new value in analytics and business intelligence.

Big Data is fueling new value in analytics and business intelligence.

¹ Richard Winter, WinterCorp Top Ten Survey, September 2005 http://www.wintercorp.com/PressReleases/tp2005_pressrelease_091405.htm

² Special Edition, Data, Data Everywhere The Economist February 2010. http://www.economist.com/node/15557443?story_id=15557443

³ Special Edition, All too much data, The Economist, February 2010 <http://www.economist.com/node/15557421>

⁴ Stacey Higginbotham, Sensor Networks Top Social Networks for Big Data, GigaOm, September 2010 <http://gigaom.com/cloud/sensor-networks-top-social-networks-for-big-data-2/>

⁵ TechCrunch.com, Twitter Seeing 6 Billion API calls a day, 70 per second, September 2010 <http://techcrunch.com/2010/09/17/twitter-seeing-6-billion-api-calls-per-day-70k-per-second/>

Getting Big Value From Big Data... Fast

Managing Big Data

Having more data should foster better decision-making. It should increase accuracy and insight. In years past BI professionals had to forego speed and deep insight because they were forced to archive or eliminate historical data due to cost considerations or limitations of their BI tools and systems. Selecting too much data for a query could relegate you to the production queue where your job could wait for hours before the data gatekeepers allowed your large report to run. Or perhaps worse, you could bring your systems to a grinding halt by attempting to include Big Data in your decision processes. These restrictions curtailed the growth of pervasive BI and forced end users to settle for “good enough” business intelligence. It also injected a control point or barrier to data that was often governed by IT. This created an adversarial relationship that negatively affected the impact of BI within the enterprise.

Many companies are no longer tolerating these types of physical or cultural restraints to their mission-critical BI environments. They have issued a mandate to corporate IT and the vendor community demanding innovation and new strategies to overcome these issues. This innovation will require new architectures and functionality along with a need to address Big Data with tools that are specifically designed to leverage it.

Necessary components to leveraging Big Data for adding value to enterprise BI include:

- **Self-Service BI** is key to growing a successful BI community. It is a paradigm that can only be driven with tools that bridge the gap between IT and the business user. Fluid, powerful and easy-to-operate User Interfaces (UIs) that support both exploration and presentation of information are key elements to enabling more business end users to embrace BI and to making better and stronger decisions. The most successful companies will add a culture of meritocracy that creates an environment where utilization of these tools rewards employees for partnering between IT and business owners to run the company better.
- **In-Memory Analytics** is a technical advancement that adds speed to analytics by accessing data that has been loaded into RAM memory. Storing data in-memory makes it easier for end users to slice and dice data and supports exploration at a faster level without the limitations often imposed by multidimensional cubes. This opens the door to analyze greater amounts of data in real-time or in an ad-hoc fashion. Eliminating the fixed Input/Output (I/O) speed of traditional disk drives creates instant value and faster decisions. Based on the specific computing environment and analytic needs of the end users the sheer volume of Big Data can create limitations for some in-memory environments. The ability to directly connect to Big Data can add flexibility, speed and support larger and more advanced workloads. The opportunity to choose from in-memory or direct data access will help users solve significant Big Data challenges.
- **Wide Data Access** supports the needs of self-service BI users who can't or don't want to get involved with the intricate aspects of data integration and ETL. This type of feature empowers the business user and relieves the IT group of unnecessary chores. End users look for opportunities to connect system information with spreadsheets on the desktop to do analysis – innovative BI platforms need to support this ability to federate data from multiple sources in an easy to understand manner

Getting Big Value From Big Data... Fast

- **Collaboration** is a growing trend among BI professionals. This is especially true with those joining the workforce today. The “Google Generation” of knowledge workers entering our workplace expects to find data fast and be able to share it with others in order to make the best business decisions. EMA criteria for successful social/collaborative business intelligence include:
 - **Access** – Diverse and varied access to data and tools
 - **Evaluate** – Leverage BI tools to gain perspective
 - **Collaborate** – Include other stakeholder and influencers
 - **Sharing** – Publish, mash-up and distribute
- **Data Blending** or data mash-ups empower end users who need to augment their analysis with multiple data sources. The ability to add new data to an analysis with a simple drag and drop interface empowers end users to ask and answer business questions without engaging IT. Traditionally users gathered disparate information and brought it to the desktop where they were forced to use Excel for analysis. This process was obviously limited in features, size, scope and speed. Data blending within your analytic environment quickens the time to decisions and maintains the ability to analyze Big Data while splicing in additional data to model new scenarios. Sophisticated data blending solutions will allow for:
 - The ability for end users to blend data sources as needed
 - Simple integration processes (e.g., drag and drop)
 - Ability to perform calculations on blended data sets
 - Integrate blended data into reports and dashboards
 - Filter across blended data sources

Getting Big Value from Big Data

The components of Big Data: size, source, type and velocity are all challenges to value for BI platforms. This is especially true of the velocity of Big Data. High-speed data or real-time data can be especially helpful in making business decisions and are critical when analyzing financial streaming data, machine or sensor data. Capturing and delivering the information so you can analyze it and compare it to other trends and historical information is difficult but highly valuable. It is at this junction that speed of data can collide with speed of decision to reduce overall BI value. If you’re able to reduce data latency, analytic latency and decision latency, business value is created in the process. (See Figure 1)

To extract that greatest level of value from your data it’s important to utilize tools and architectures that support the reduction of data latency, analytic latency and decision latency. Self-service business intelligence, in-memory analytics, wide data access and collaboration are all contributing elements to affecting this reduction.

Getting Big Value From Big Data... Fast

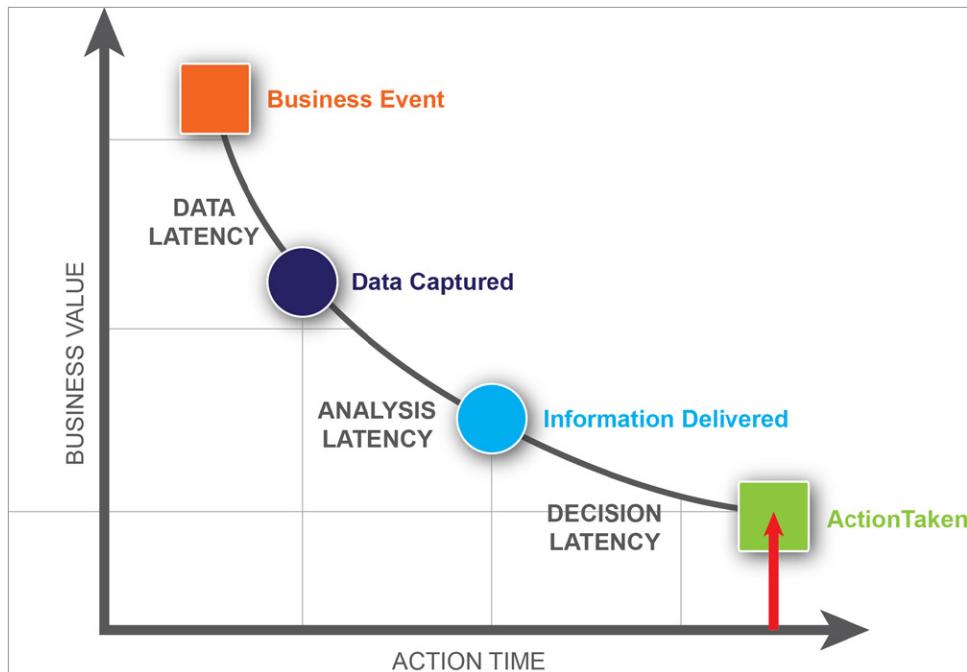


Figure 1 (Based on concept developed by Richard Hackathorn, Bolder Technology⁶)

Tableau Software Technology

Tableau Software, based in Seattle, Washington, started out as a Stanford University project. The company has quickly established a leading position in the BI space and today is serving over 50,000 users worldwide. Their dedication to powerful and easy-to-use functions has provided an alternative path for BI professionals who want to create fast ROI and enable a diverse user community across the enterprise. Tableau combines brilliant data visualizations with the ability to explore data.

Tableau data visualizations provide a starting point to quickly engage with business data. They support a powerful layer of exploration that allows the end user to investigate and unearth valuable “ah-ha” moments that cannot be identified with less featured BI platforms. Tableau’s in-memory technology enables users to achieve speed and flexibility in analysis by optimizing slower data, while its option to connect directly to databases means that customers who have invested in fast data infrastructure can leverage that speed as part of their BI system.

Tableau supports connections to more than 20 data sources, from Microsoft Excel PowerPivot data to Teradata, Vertica, SQL Server and Oracle. Access to diverse data sets open the door for true BI exploration and put much of the power in the hands of the end user. Tableau adds more value to the environment with data blending features that allow accessing data from multiple sources, finding fields in common and working with the blended data in one view. Tableau’s most recent release, version 6.0, contains 60+ upgrades that continue to push Tableau forward in the market.

⁶ Dr. Richard Hackathorn, Real-Time to Value Research Bolder Technology 2009

Getting Big Value From Big Data... Fast

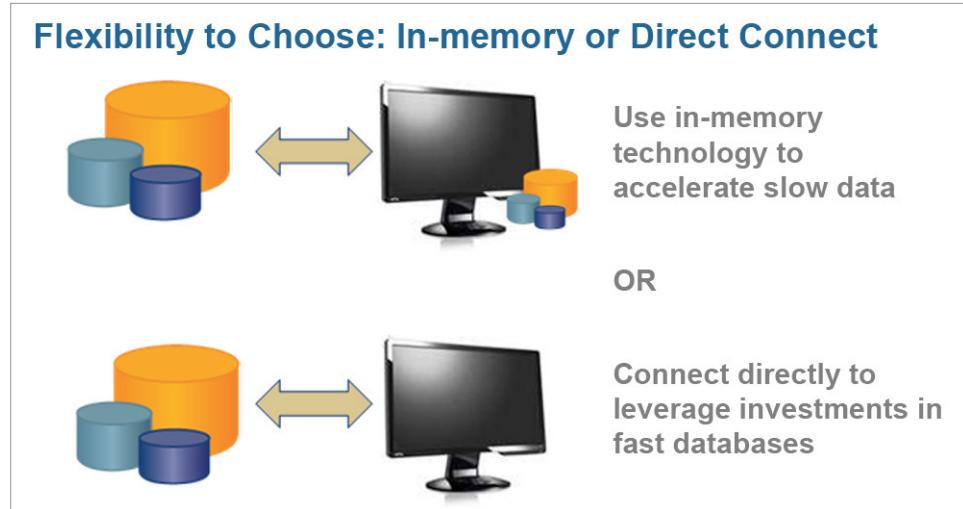


Figure 2

EMA Perspective

Tableau represents a new breed of business intelligence and analytic platforms that is designed to leverage today's growing Big Data challenges along with empowering the end-user community with fast, easy-to-use environments that couple both information presentation features with data exploration. It is the best of both worlds and delivers flexibility that serves the needs of business users and power users alike. In the end these new-style BI platforms reduce IT dependency, save time and money. EMA sees Tableau as a leader in this space and expects that through continued investment and innovation that they will maintain a leadership role in this segment of business intelligence platforms.

. EMA sees Tableau as a leader in this space and expects that through continued investment and innovation that they will maintain a leadership role in this segment of business intelligence platforms.

Customer Perspective

The success of Tableau Software is best reflected in its customers. A recent project with Yahoo! designed to optimize ad placement on Yahoo! media properties has blossomed into 400 Yahoo! employees worldwide working with Tableau Server, 100 of whom are using the system to explore data with Tableau Desktop. The Tableau platform sits over what Yahoo! claims is the largest multi-dimensional database in the world, a 12-terabyte MOLAP cube that loads four billion new records daily. The Yahoo! team uses Tableau to optimize ad campaigns on behalf of its advertisers selecting the optimal ads to present and publisher to target with the ads. The ability to include more data in their analysis has proven valuable for Yahoo!. Tableau is able to deliver ad size, audience segments, geography, age, gender and other dimensions that were missing from prior analysis. Yahoo! leverages Tableau to serve predesigned dashboards with full slice-and-dice and drilldown capabilities while giving team members the ability to roam free and explore the data environment to discover new ideas and insights. The ability for Yahoo! team members to execute ad-hoc style queries on big data is extremely valuable and supports these line of business users with the ability to discover and explore Big Data in

Getting Big Value From Big Data... Fast

a way they were unable to in the past, reducing the time to decision gap and making Yahoo! a more agile company. Yahoo! reports a quick learning curve, excellent adoption and ease of use, and Big Data scalability as factors to the projects success.

Another example of Big Data success is Playdom, a social gaming firm owned by The Walt Disney Company. Playdom delivers online games hosted on social networking platforms such as Facebook and MySpace. Over eight million people play these games every day producing over a billion rows of information and creating upwards of two terabytes of data. Playdom was challenged by the need to examine information in real-time to better understand the engagement and retention levels of their gamer community. Playdom utilizes Tableau Server along with Vertica Systems, a column store data warehousing solution, to serve the needs of over 80 product managers. Speed and Big Data play a key role in Playdom's analytic strategy. The combination of Tableau and Vertica enables them to query a greater number of dimensions and size to gain insight into the drivers behind game adoption and player retention. Perhaps the largest value the system delivers to Playdom is the ability to iterate quickly. Previously they were challenged by queries that took over an hour to execute causing business decisions to take days. Tableau is providing similar analysis with more data in 10 to 15 seconds allowing Playdom to explore and run iterative analysis to make better decisions in real-time.

Final Thoughts on Big Data

Big data is here to stay and is adding immense value as well as challenges to the BI and analytics space. Speed, flexibility and scalability all play a significant role in successfully leveraging Big Data. A couple key points to remember:

- Reducing data latency can create more value because it reduces the overall “action time” from business event to action taken. Make sure that as your data volumes expand, you focus on systems that reduce data latency.
- Big data needs more than just big databases to make sense of it. Seek systems that bridge the gap between IT and users.
- Despite the ever-growing volumes, users still need data to be fast. Don’t expect your traditional data warehouse to always solve that problem. Look for applications that can leverage the big data data warehouse but also provide user-friendly, in-memory capabilities for speed.

About Tableau Software

Tableau Software is the leading provider of fast analytics and data visualization software. Tableau's award-winning business intelligence applications can be downloaded at <http://www.tableausoftware.com/trial>. They enable anyone to easily create and share interactive data visualizations, dashboards and analytics, and can scale to organizations of any size or reach. For more information, please visit <http://www.tableausoftware.com> or visit the company blog: <http://www.tableausoftware.com/about/blog>.

About Enterprise Management Associates, Inc.

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst firm that provides deep insight across the full spectrum of IT and data management technologies. EMA analysts leverage a unique combination of practical experience, insight into industry best practices, and in-depth knowledge of current and planned vendor solutions to help its clients achieve their goals. Learn more about EMA research, analysis, and consulting services for enterprise IT professionals, lines of business users, and IT vendors at www.enterprisemanagement.com or follow [EMA on Twitter](#).

This report in whole or in part may not be duplicated, reproduced, stored in a retrieval system or retransmitted without prior written permission of Enterprise Management Associates, Inc. All opinions and estimates herein constitute our judgement as of this date and are subject to change without notice. Product names mentioned herein may be trademarks and/or registered trademarks of their respective companies. "EMA" and "Enterprise Management Associates" are trademarks of Enterprise Management Associates, Inc. in the United States and other countries.

©2011 Enterprise Management Associates, Inc. All Rights Reserved. EMA™, ENTERPRISE MANAGEMENT ASSOCIATES®, and the mobius symbol are registered trademarks or common-law trademarks of Enterprise Management Associates, Inc.

Corporate Headquarters:

5777 Central Avenue, Suite 105
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com

