

White Paper

Big Data Visualization: Turning Big Data Into Big Insights

The Rise of Visualization-based Data Discovery Tools

Why You Should Read This Document

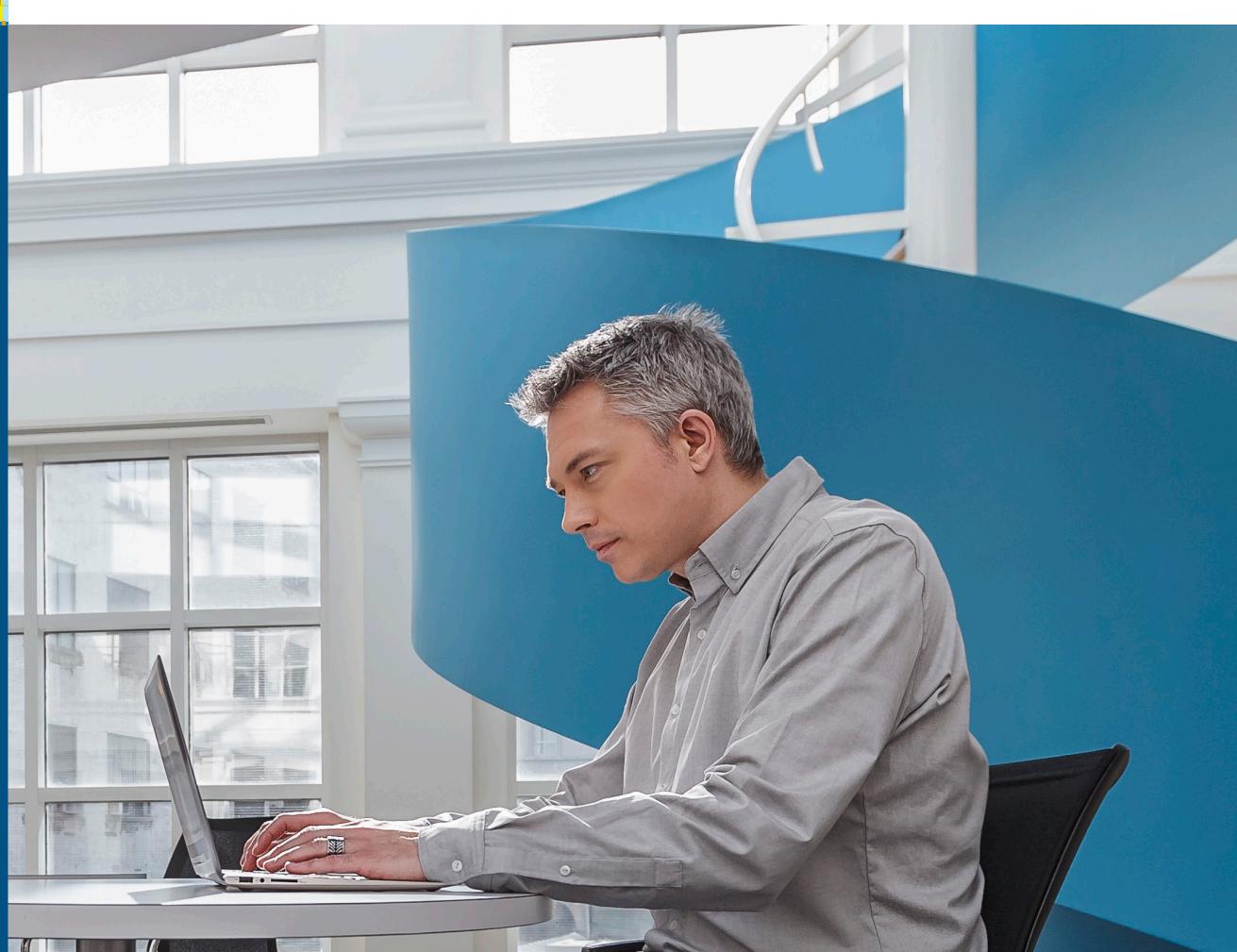
This white paper provides valuable information about visualization-based data discovery tools and how they can help IT decision-makers derive more value from big data. Topics include:

- An overview of the IT landscape and the challenges that are leading more businesses to look for alternatives to traditional business intelligence tools
- A description of the features and benefits of visualization-based data discovery tools
- Guidance and suggestions on data governance, and ways to protect the quality of big data while facilitating self-service business intelligence
- Several usage examples of visualization-based data discovery tools from TIBCO* Software, the world's second-largest data discovery vendor

White Paper

Turning Big Data Into Big Insights

The Rise of Visualization-based Data Discovery Tools



Contents

- [3 The Trend Toward Visualization-based Data Discovery Tools](#)
- [4 The Struggle to Make Meaning Out of Big Data](#)
- [5 Deriving Real Value from Big Data](#)
- [7 Protecting Data Quality](#)
- [8 Usage Examples from TIBCO Software](#)

The Trend Toward Visualization-based Data Discovery Tools

Big data is creating unprecedented opportunities for businesses to achieve deeper, faster insights that can strengthen decision-making, improve the customer experience, and accelerate the pace of innovation. But today, most big data yields neither meaning nor value. Businesses are so overwhelmed by the amount and variety of data cascading into and through their operations that they struggle just to store the data—much less analyze, interpret, and present it in meaningful ways.

For help, businesses are increasingly turning to visualization-based data discovery tools, leading Gartner to estimate a 30 percent compound annual growth rate through 2015.¹ The tools promote self-service business intelligence (BI), enabling a multitude of users to easily integrate or “mash up” data from a wide range of sources—clickstreams, social media, log files, videos, and more. With the aid of high-powered desktops and mobile computing devices such as the Ultrabook™, users can perform real-time, predictive analyses, and showcase the results in compelling, interactive, and easily understood visual formats.

The trend toward visualization-based data discovery tools is worth exploring by any business that seeks to derive more value from big data. The potential business benefits are immense, and data governance best practices can be used to help ensure a safe transition. As demonstrated by three usage examples from TIBCO Software*, the world’s second-largest data discovery vendor, real-world applications of visualization-based data discovery tools are already delivering greater customer and market insights to businesses around the world.

“By visualizing information, we turn it into a landscape that you can explore with your eyes, a sort of information map. And when you’re lost in information, an information map is kind of useful.”

*-David McCandless,
author, data journalist, and
information designer*

The Struggle to Make Meaning Out of Big Data

Data analytics and visualization are not new. For decades, businesses have collected data, analyzed it using a variety of BI tools, and generated reports. The process may take weeks or months, but eventually a few highly trained data analysts are able to pull the necessary figures from their dashboards and issue static, rearview reports to executives and other employees.

Businesses are finding that this traditional reporting process does not work nearly as well for big data, and certainly is not sufficient to capture the potential value that big data represents. The primary challenges stem from what are commonly termed the “three Vs” of big data: volume, variety, and velocity. Most traditional reporting and data mining tools cannot handle the vast volume of big data—although the variety and velocity of the data often present even greater challenges. .

Another key challenge in analyzing big data relates to its velocity. The rapid generation of big data can lead to significant business insights and predictions, but only if real-time data can be analyzed quickly—in hours rather than weeks or months. Reducing the latency from data capture to action is absolutely vital. Today, however, Intel’s IT Manager survey found that only about half of IT managers perform data analytics in real time, while the other half continue to rely on batch processing that fails to capture the immediacy of big data.⁴

A final challenge driving the rise of visualization-based data discovery tools is the increasing availability of mobile devices. Businesses that continue to rely on centralized creation of reports by a few highly trained experts are missing an opportunity to adopt a faster, more cost-effective, and more democratized BI model that takes advantage of the intersection of big data and the mobile workforce to speed insights and improve collaboration.

Big data includes three types of data—structured, semistructured, and unstructured—and Intel’s IT Manager Survey of 200 IT professionals found that four of the top five data sources for IT managers today are semistructured or unstructured². Many businesses are simply unable to analyze these emerging forms of data, which include everything from e-mails, photos, and social media to videos, voice, and sensor data. In fact, an IBM survey of more than 1,100 business and IT professionals found that fewer than 26 percent of respondents who had active big data efforts could analyze extremely unstructured data such as voice and video, and just 35 percent could analyze streaming data.³

Key Results from Intel’s IT Manager Survey

- *33% of companies surveyed are working with very large amounts of data (500 TB or more)*
- *84% of IT managers are analyzing unstructured data.*
- *44% of those who are not analyzing unstructured data expect to do so in the next 12 to 18 months*
- *By 2015, IT managers expect that 63% of all analytics will be done in real time*
- *Of seven possibilities, IT managers indicated that they would find the most value in receiving help deploying cost-effective data visualization methods*

Deriving Real Value from Big Data

While Apache* Hadoop* and other technologies are emerging to support back-end concerns such as storage and processing, visualization-based data discovery tools focus on the front end of big data—on helping businesses explore the data more easily and understand it more fully.

Visualization-based data discovery tools allow business users to mash up disparate data sources to create custom analytical views with flexibility and ease of use that simply didn't exist before. Advanced analytics are integrated in the tools to support creation of interactive, animated graphics on desktops, as well as on powerful mobile devices such as the Ultrabook™ and laptops powered by Intel® Core™ processors. End users can view the graphics on the same devices, or on even smaller mobile devices such as tablets or, in limited cases, smartphones.

Because of their ease of use and intuitive interfaces, visualization-based data discovery tools have a democratizing effect on businesses. Data analysis and visualization—formerly the province of only a limited handful of highly trained data analysts—can be accomplished by a multitude of users with minimal training. Moving toward a self-service model for BI can reduce costs and enable IT to spend more time focusing on business-building innovations and complex data challenges.

Self-service BI also enables businesses to take advantage of increasingly mobile workforces. For instance, remote and on-site members of a product development team can easily view and share visualizations that explore potential product defects or customer preferences. The bring-your-own-device (BYOD) trend means that these users can use their own mobile devices to easily explore the data, discover trends and patterns, and communicate their findings to fellow team members and other audiences.

Key Features of Visualization-based Data Discovery Tools

- Enable real-time data analysis
- Support real-time creation of dynamic, interactive presentations and reports
- Allow end users to interact with data, often on mobile devices
- Hold data in-memory, where it is accessible to multiple users
- Allow users to share and collaborate securely

Additional Features to Look For

- Ability to visualize and explore data in-database as well as in-memory
- Governance dashboard that displays user activity and data lineage
- In-memory data compression to enable handling of large datasets without driving up hardware costs
- Touch optimization for use with touch-enabled mobile devices such as the Ultrabook™

Addressing the Three Vs

Visualization-based data discovery tools take the challenges presented by the "three Vs" of big data and turn them into opportunities for growth.

Volume

Unlike most traditional BI systems, visualization-based data discovery tools are designed to work with an immense number of datasets, so businesses can turn their attention from simply managing the deluge of data to gaining rich insights. From visualizing national marketing campaigns to mining and presenting sales data, the tools enable businesses to derive meaning from large, and growing, volumes of data.

Variety

Visualization-based data discovery tools are designed to mash up, or combine, as many data sources as needed. That means businesses can derive more meaning from structured data, as well as semistructured and unstructured data sources such as social media and sensor data. Using interactive bubble charts, 3-D data landscapes, treemaps, boxplots, heatmaps, word clouds, and many other types of graphics, businesses can view, interpret, and interact with complex data from a multitude of sources.

Velocity

With visualization-based data discovery tools, businesses can replace batch processing with real-time processing of continually updated data streams. The tools also support the democratization of data discovery, so more people can access real-time data sources such as clickstreams, and analyze and view the data without having to wait for reports.

The 4th V of Big Data



When businesses address the three Vs in parallel, they achieve the fourth V: **Value**. Visualization-based data discovery tools don't just enable users to create attractive infographics and heatmaps. They create business value by enabling more workers to gain more insights from more data. Instead of waiting weeks or months for static reports, employees can analyze and visualize real-time data on their own. They can also collaborate with co-workers using online, interactive graphics to generate new ideas and identify previously unseen trends.

All these benefits eventually translate to a better bottom line—with stronger sales, better customer experiences, and market-leading innovations.

Ultrabook™ Convertible

One of the best and most flexible mobile devices for visualization-based data discovery is the Ultrabook™ convertible, which combines the full PC performance needed to analyze and visualize large datasets with the convenience and responsiveness of a tablet.

Powered by Intel® Core™ processors, Ultrabook™ convertibles are less than an inch thick and wake up in a flash⁵ to provide fast access to applications and files. Many models are also touchscreen enabled so users can explore graphs and charts quickly and intuitively. To keep the devices safe, unique anti-theft and identity protection security technologies are built in to the processor.^{6,7}

Protecting Data Quality

Data security and governance have always been part of BI, but big data introduces added legal, ethical, and regulatory issues. Visualization-based data discovery tools further those concerns, particularly in the area of data quality.

The risk to data quality stems from one of the great benefits of visualization-based data discovery tools: their ease of use. The tools facilitate self-service BI, enabling more users to perform advanced analyses. That in turn leads to a legitimate risk of creating not only more data errors but also what Gartner terms "disconnected 'rogue' data discovery islands."⁸

The risks can be greatly reduced by developing safe frameworks within which businesses can seize the immense opportunities presented by visualization-based data discovery tools.

Data governance best practices

Intel's IT Manager survey found that one of the three top challenges for IT managers is the issue of data governance and the need to have a policy for defining how data will be stored, analyzed, and accessed.⁹

No single set of governance best practices has emerged specifically for integrating visualization-based data discovery into businesses, but attendees at a Gartner BI Summit recommended that businesses:

- **Require the certification of data sources from which reports or dashboards can be generated**
- **Store and analyze users' record updates to identify potential user errors**
- **Use data profiling techniques to publish quality rankings**
- **Assess which user or data steward has created or maintained the best data quality metrics¹⁰**

Instead of trying to prevent the democratization of data discovery tools, Gartner suggests that businesses consider creating IT-sanctioned "sandboxes" where users can safely analyze data and share results. This tighter collaboration between IT and users has the potential to prevent data discovery tools from becoming "rogue" or isolated data-mart solutions.¹¹

To support self-service BI, Gartner further suggests that businesses:

- **Create organizational structures that blend IT and business skills, and strike a balance between centralized and decentralized BI delivery**
- **Invest in consumerization technologies such as mobile devices, interactive visualization tools, and search applications to increase user adoption**
- **Empower users to create their own analytical views, but also provide a way to certify this content for internal distribution¹²**

With a proper framework in place to control and oversee the democratization of data, businesses can maintain confidence in the quality of their data and reports. At the same time, they can take full advantage of the many benefits of visualization-based data discovery tools.

A different approach to collaboration

TIBCO takes a unique approach to self-service BI. While others in the industry emphasize the full democratization of information, with databases open to all end users, TIBCO encourages a slightly more conservative model that is similar to the "sandbox" approach suggested by Gartner.

TIBCO Senior Director of Analytics Michael O'Connell said, "I'm all for self-service, but our goal is to provide a more guided self-service approach. We believe that some level of control is essential, especially for larger businesses where governance issues are a concern."

In TIBCO's model for BI collaboration, trained business analysts—sometimes called "data scientists"—can access databases and create templates that are then saved to the Spotfire library. End users with library credentials can easily access the templates, which become building blocks for their own self-service analyses. The key is that the templates provide a framework for the analyses—enabling end

users to drill down, filter data, and change views, for instance, while preserving the analyst framework and workflow guidance.

When end users have insights, ideas, or hypotheses, they can create "social bookmarks" that enable them to share their thoughts with select groups—including peers, advisers, and customers. Works in progress can be marked with "private bookmarks" to keep them out of group discussions while keeping track of analyses and insights for future reference.

"We're a little more conservative in wanting some degree of governance and guidance rather than unlimited end-user access to data, but we still believe in the value of democratization," said O'Connell. "Authors can create templates quickly, and end users have an interactive environment in which they can generate critical insights to address a broad class of business problems."

Spotfire features

Spotfire 5.0, the most recent product release from TIBCO, shows the growth and potential of visualization-based data discovery tools. Spotfire has long enabled users to mash up huge data sets in-memory, and the new release features a much-enhanced in-memory analytics engine that takes better advantage of high-capacity, multi-core servers such as those based on Intel® Xeon® processors.

For applications involving terabytes of data, Spotfire 5.0 now supports in-database and on-demand analysis in addition to in-memory data discovery. Users can perform analytics within powerful database platforms such as Teradata*, Oracle*, and Microsoft* SQL Server, where large data sets reside, so less time is spent extracting and moving data. Results are then returned to Spotfire, enabling users to drill down, up, and sideways, from in-database aggregations to detailed on-demand and in-memory analytics.

Spotfire 5.0 also supports analytics based on the popular R programming language. TIBCO Enterprise Runtime for R embeds a TIBCO-proprietary R engine into the Spotfire product stack, both client-side and server-side. Running R within Spotfire makes it possible to deploy powerful R-based applications to thousands of users, addressing a valuable set of business use cases with prediction, forecasting, optimization, machine learning, and simulation—all inside the data discovery environment.

Advantages of Intel® Core™ Processors

As TIBCO's senior director of analytics, O'Connell spends a great deal of time on his own computer, creating and testing visualizations and collaborating with team members. The laptop he relies on is powered by a 3rd generation Intel® Core™ i7 vPro™ processor.¹³

The Intel Core i7 vPro processor provides top-of-the-line processing speed for multitasking and multimedia tasks. In addition, it features advanced graphic features such as Intel® Clear Video HD Technology¹⁴ for sharper, smoother images that bring TIBCO Spotfire visuals to life.

Intel® vPro™ technology provides an additional set of security and manageability capabilities. The features are built into the hardware to provide an added layer of protection against difficult-to-detect, penetrating rootkits, malware, and other threats. The manageability features also enable remote, anytime access to O'Connell's PC for monitoring and maintenance, which helps increase uptime.

Usage Examples from TIBCO Software

Chevron's iRAVE Data Discovery Solution

Every day, Chevron's asset teams review the company's oil and gas fields for production shortfalls and mitigation efforts. In most fields, however, there are too many wells to examine them all individually. To address this challenge, Chevron used TIBCO Spotfire to develop iRAVE (Integrated Reservoir Analysis and Visualization Environment). iRAVE integrates and exposes relevant data sources from the oil fields, and presents the data in a user-friendly visual analytics environment.

iRAVE has been tested and deployed to global Chevron assets. In each deployment, iRAVE has resulted in 10-fold ROI, and a payback within the first three months of use. The solution enables asset teams to understand operator issues and optimize the assets while substantially improving team productivity far more than ever anticipated. Another key advantage is the adaptability and IT support for Spotfire across multiple functional areas, which ensures that the platform and data are protected.

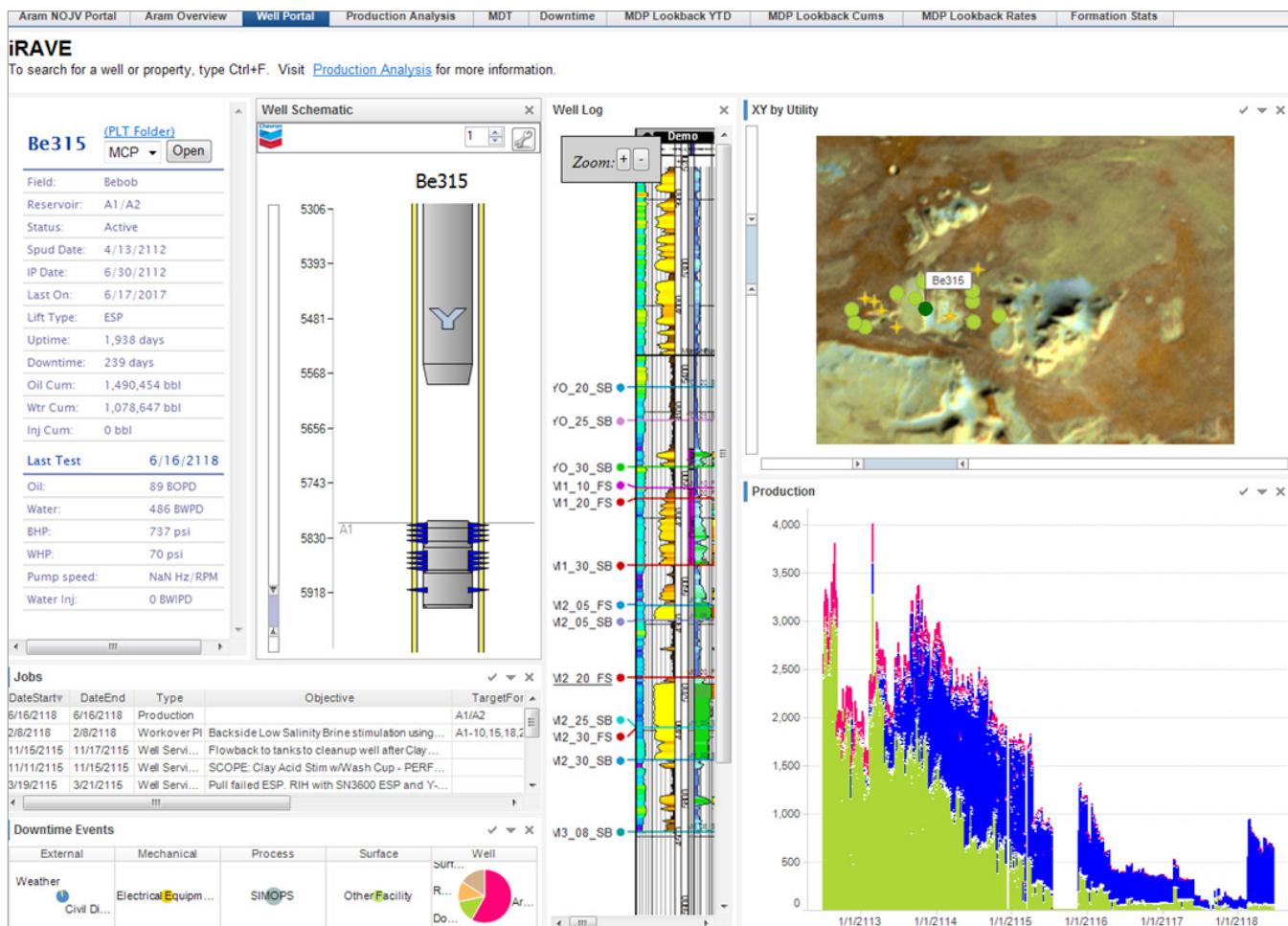


Figure 2. Chevron's iRAVE application for well production surveillance and optimization.

Pfizer's Clinical Trials Data Visualization

On any given day, Pfizer is conducting hundreds of clinical trials around the world. The company uses TIBCO Spotfire to visualize the multitude of data collected on more than 500 of these trials. Specialized Spotfire visualizations enable Pfizer to explore the clinical data for safety, efficacy, and operations tracking.

Pfizer staff in the U.S., Europe, and the Asia-Pacific regions log in to the Spotfire web-based data discovery application, where they can easily explore clinical trial and patient-level data. Pfizer data scientists, working with TIBCO, designed the workflows, and the end-user community is able to interactively review the multitude of clinical data across the trial population and at the individual patient level.

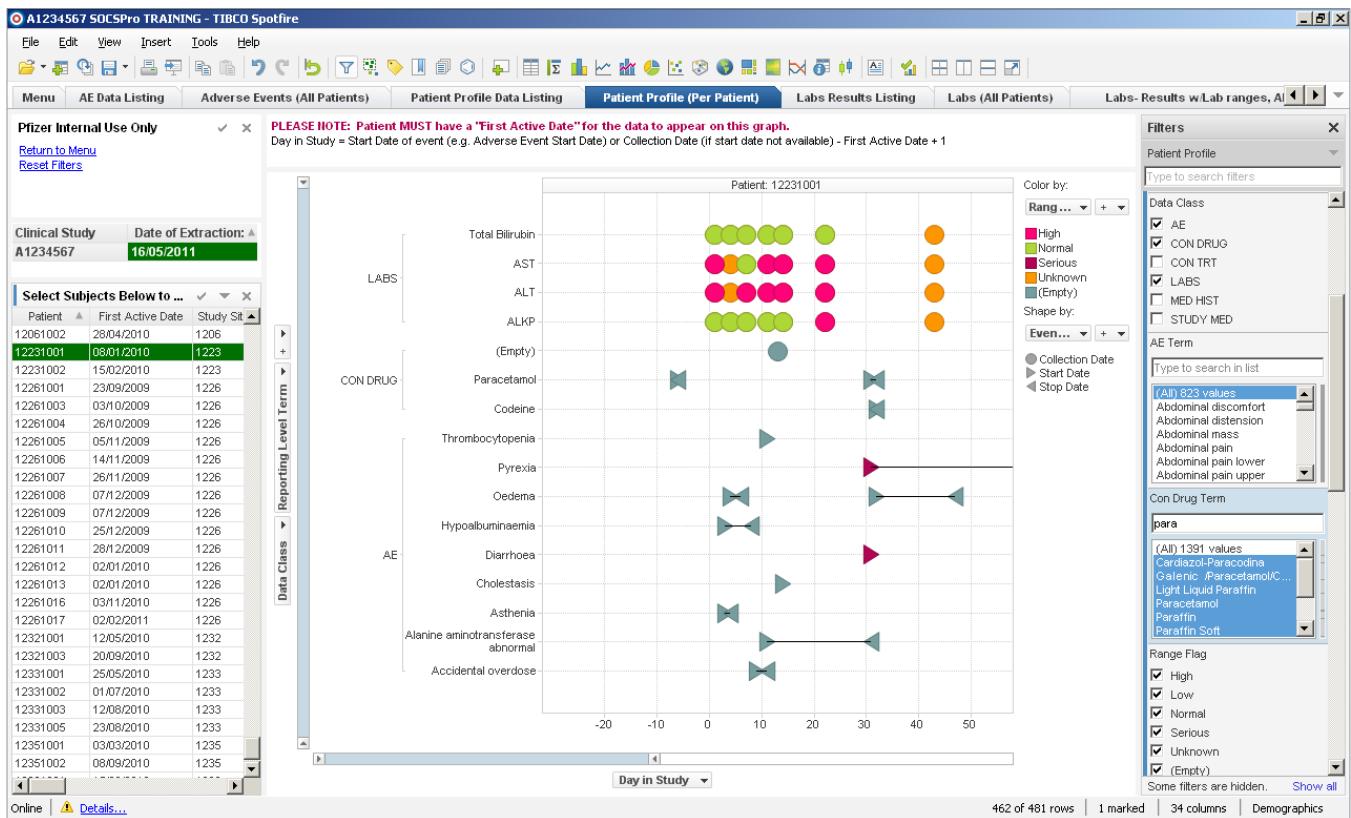


Figure 3. Pfizer uses a highly interactive patient profile in Spotfire to relate patient-level information and assess patient safety.

What's Next?

The strong and growing trend toward visualization-based data discovery tools is expected to continue in the years to come, as more businesses seek better, more cost-effective ways to derive meaning from their big data.

Visualization-based data discovery tools provide an immense opportunity—not only to manage the growing volume, variety, and velocity of new and existing data but also to turn that data into value. Using high-powered desktops and mobile devices, and with the help of firm data governance practices and vendors such as TIBCO Software, businesses can better understand operations, customers, and the marketplace. And those faster, deeper insights can lead to greater business agility and a significant competitive advantage moving forward.

Learn More

To learn more about big data and visualization-based data discovery tools, visit the IT Center at intel.com/bigdata.

Endnotes

1. Dan Sommer, Rita L. Sallam, James Richardson, "Emerging technology analysis: Visualization-based data discovery tools," June 17, 2011.
2. Intel IT Center, "Intel's IT Manager survey on how organizations are using big data," August 2012.
3. IBM Institute for Business Value, in collaboration with SAID Business School at the University of Oxford. "Analytics: The real-world use of big data." 2012.
4. Intel IT Center, "Intel's IT Manager survey on how organizations are using big data," August 2012.
5. SRequires a select Intel® processor, Intel® software and BIOS update, and Intel® Solid-State Drive (SSD). Depending on system configuration, your results may vary. Contact your system manufacturer for more information.
6. No system can provide absolute security under all conditions. Requires an enabled chipset, BIOS, firmware, software, and a subscription with a capable service provider. Consult your system manufacturer and service provider for availability and functionality. Service may not be available in all countries. Intel assumes no liability for lost or stolen data and/or systems or any other damages resulting thereof. For more information, visit <http://www.intel.com/go/anti-theft>.
7. No system can provide absolute security under all conditions. Requires an Intel® Identity Protection Technology-enabled system, including a 3rd or 4th gen Intel® Core™ processor-enabled chipset, firmware, software, and participating website. Consult your system manufacturer. Intel assumes no liability for lost or stolen data and/or systems or any resulting damages. For more information, visit <http://ipt.intel.com>.
8. Dan Sommer, Rita L. Sallam, James Richardson, "Emerging technology analysis: Visualization-based data discovery tools," June 17, 2011.
9. Intel IT Center, "Intel's IT Manager survey on how organizations are using big data," August 2012.
10. Andreas Bitterer, Bill Gassman, James Richardson, "2012 BI Summit hot topics: Data quality, big data, strategy and visualization," Sept. 26, 2012.
11. Dan Sommer, Rita L. Sallam, James Richardson, "Emerging technology analysis: Visualization-based data discovery tools," June 17, 2011.
12. Kurt Schiebel, "How to deliver self-service business intelligence," May 10, 2012.
13. Intel® vPro™ Technology is sophisticated and requires setup and activation. Availability of features and results will depend upon the setup and configuration of your hardware, software and IT environment. To learn more visit: <http://www.intel.com/technology/vpro>.
14. Built-in visual features are not enabled on all PCs and optimized software may be required. Check with your system manufacturer. Learn more at <http://www.intel.com/go/biv>.
15. http://www.pg.com/en_US/downloads/innovation/factsheet_BusinessSphere.pdf
16. www.informationweek.com/global-cio/interviews/pg-ceo-shares-3-steps-to-analytics-drive/240148065

Share with Colleagues [f](#) [in](#) [t](#) [e-mail](#)

This paper is for informational purposes only. THIS DOCUMENT IS PROVIDED 'AS-IS' WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL SPECIFICATION, OR SAMPLE. Intel disclaims all liability, including liability for infringement of any property rights, relating to use of this information. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted herein.

Copyright © 2013 Intel Corporation. All rights reserved.

Intel, the Intel logo, Intel Sponsors of Tomorrow., the Intel Sponsors of Tomorrow logo., Xeon, Core, vPro, and Ultrabook are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

328729-001EN

