



The Buyer's Guide to Business Intelligence Software: Market Overview

2016 edition

Curated from End-User Reviews on:

trustradius[™]

First published June 2016

© 2017 TrustRadius. All rights reserved. This publication has been licensed by TechTarget. Reproduction or sharing of this publication in any form without prior written permission is strictly prohibited.



Table of Contents

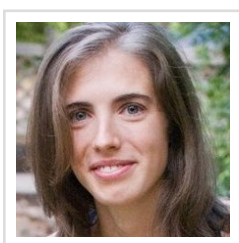
About This Report.....	3
Types of Business Intelligence Tools	5
Recent Trends in the Business Intelligence Market.....	13
How to Buy a BI Solution.....	23

About This Report

Why read this report?

TrustRadius is excited to publish our second edition of the Buyer's Guide to Business Intelligence Software. We hope this report will help you select the BI tool that best fits your needs, by providing you with market context and expert tips about how to approach your evaluations.

In addition to reading and distilling more than 2,250 ratings and in-depth reviews of 22 leading BI software products on TrustRadius, we spoke with vendors and experts to offer differing viewpoints on the direction of the BI market in 2016. The result is a reliable, independent analysis, uninfluenced by vendors or anyone else with an agenda.



Megan Headley
Research Director
TrustRadius

About This Business Intelligence Guide: A Note from the Author

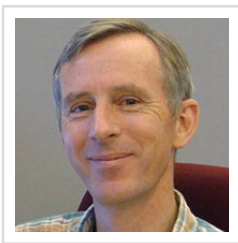
The first TrustRadius Buyer's Guide to Business Intelligence was published in November 2014. That guide segmented the market into different product types with different advantages and disadvantages, with some types being more suitable for specific use cases.

In the first edition, we covered two principal product groupings: traditional full-stack platforms, and more agile data discovery and visualization tools. Dashboards and predictive analytics were the remaining adjacent product types. While these two major segments still contain sets of products that are quite unlike each other, the emphasis on speed agility in today's business climate has brought the two segments closer towards each other as they both endeavor to provide data for rapid and informed decision-making. Indeed many larger organizations use products from each major segment to solve different and specific problems. However, the impetus towards decentralized and agile decision-making has rendered the strict division of tools less clear cut than it used to be.

In this guide you'll find:

- » A discussion of the various types of business intelligence products
- » High-level summary of buyer feedback on 22 BI tools
- » Recent trends in the business intelligence software market
- » Tips from experts on how to buy a BI solution

The companion piece, *The Buyer's Guide to Business Intelligence Software: User Perspectives on BI Products*, provides TrustMaps showing the Top Rated tools for Small Businesses, Mid-Sized Companies, and Enterprises, as well as detailed profiles of 22 business intelligence tools, customer demographics and pros & cons as cited in 2,250+ authenticated end-user reviews and ratings.



Alan Cooke
Research Manager
TrustRadius

ABOUT TRUSTRADIUS

TrustRadius is the most trusted review site for business technology, serving both buyers and vendors. We help buyers make better product decisions based on unbiased and insightful reviews. We also help vendors scale and harness in-depth reviews to accelerate sales. Unlike software directories, TrustRadius collects reviews that are structured and substantive, averaging more than 400 words. Every reviewer on TrustRadius is authenticated and every review vetted by our Research Team before publication. Founded by successful entrepreneurs and backed by the Mayfield Fund, TrustRadius is bringing transparency and efficiency to the \$3.7 trillion business technology market.

To learn more, visit www.trustradius.com.

Types of Business Intelligence Tools

The Data Management Problem that BI solves

Companies collect large quantities of operations data as a by-product of doing business. Huge quantities of data are stored in finance, procurement, sales, marketing systems and multiple other data repositories. Being able to analyze and understand this data is extremely important to running the business. For example, Enterprise Resource Planning (ERP) systems typically contain data concerning the supply chain and inventory levels in addition to financial data. HR systems contain all employee records including demographic data, salary level, and performance reviews. Customer Relationship Management (CRM) systems contain customer, sales pipeline, forecasting and sometimes customer support case data.

The problem is that all this operational data is typically not accessible in one place for analysis in order to make decisions and provide strategic guidance to the business as a whole. For example, inventory data from an ERP system could be combined with sales forecasting information to understand how to optimize inventories in response to demand. This is the problem that business intelligence systems were designed to solve.

How Business Intelligence Tools Work

Traditional business intelligence solutions solve this problem by putting data into a common store called a warehouse. The data is then normalized - removing redundancy and duplication - making it easier to run queries and retrieve data for reporting. Newer data discovery and visualization platforms solve the problem differently, by either connecting directly to the various data sources, or storing data in-memory for analysis and visualization. There are many different types of business intelligence technology, not all of which depend on the business warehouse paradigm. Many new approaches have emerged, and the following sections describe some of the major classes of business intelligence technology.

On-premise Full Stack Products

On-premise full stack BI solutions have been around the longest and are now being eclipsed by newer, more flexible technology. However, these tools still have a very large installed base, and are still very effective for managing structured data from many sources and structuring it for standard reporting across the enterprise. They have a number of key components, although every solution does not necessarily have each component of the stack:

- » **Data warehouse:** A relational database designed specifically for data analysis instead of standard transactional processing. It acts as the conduit between operational data stores and the gaining of insight based on composite data. Slices of data from the warehouse—usually summary data for a single department like sales or finance—are stored in a “data mart” for quicker access.
- » **Extract, Transform, Load (ETL):** The first important task is to extract the data from the various data sources and load it into a data warehouse where it is normalized (organized into tables while cleaning the data and removing redundancy and inconsistencies). Once it has been appropriately structured it is available for querying and analysis.
- » **OLAP or ad-hoc query tools:** OLAP (Online Analytical Processing), and its close cousin ROLAP—(Relational) Online Analytical Processing, is a technology that allows users to query data across multiple dimensions, for building standard reports or for enabling users to ask a specific business question.
- » **Presentation layer:** Dashboards, scorecards and reports presenting the data to users in a visually appealing way that is easy to understand.

These tools are useful for organizations that wish to deliver relatively stable operational reports in a consistent format to front-line staff across the organization to help them monitor their progress or understand where performance is lagging. The advantage of this kind of enterprise reporting capability is the consistency of the data sets being used across the entire organization, which makes it easy to create alignment. It is notoriously difficult to achieve alignment if there is no common agreement about the accuracy of data, and stakeholders have different sets of data showing contradictory information. This is typically what people mean when they refer to a “single source of the truth”.

However, on-premise first, full-stack BI systems are difficult to build and implement, expensive, and often difficult to learn and use. They also lack flexibility and are difficult to change once they have been built. It has been relatively common in recent years for publications and analysts to bemoan the high failure rates for BI projects, and full-stack deployments are often the culprits. Implementation times for these tools can be long, because setting up the data warehouse and creating the schema are inherently IT-intensive, complex tasks. When they are finally up and running, the ROI can be low, often because of usability problems.

However, it should be pointed out that this does not have to be so. In recent years, a new category of data warehouse automation tools has emerged to mitigate these problems. Products like TimeXtender, Kalido, WhereScape, and Attunity go some way to making data warehouse creation and maintenance a far more agile and collaborative experience. These tools are capable of automating the creation of a data warehouse schema, indexes, cubes, etc. They can also create business metadata for specific business intelligence tools. In this way, they can dramatically simplify and speed up both data warehouse development and subsequent maintenance.

Additionally, not all tools in this category are legacy tools. There are more modern approaches to providing end-to-end capabilities using newer technology. A good example of this is Sisense, which uses a more flexible version of OLAP cubes, called “elasticubes”, and leverages data storage provided by the chip set to eliminate some of the speed limitations of disk storage. This approach yields significant speed increases of more than 50x over the competition.

Full-stack BI tools built on a data warehouse can still provide immense value to larger organizations with the resources to deploy and manage them, and the deep pockets required to invest in them.

Example Products

- » [IBM Cognos](#)
- » [Microsoft BI](#)
- » [MicroStrategy](#)
- » [SAP Business Objects](#)
- » [Sisense](#)

Best Fit For

- » Organizations whose primary need is for alignment and consistency of data across a very large organization and the provision of accurate reports to line of business managers and operational employees. These tools provide “a single version of the truth” as a basis for decision making across an entire enterprise.
- » Organizations with access to a highly skilled IT division, which includes ETL developers, report developers, data architects, data administrators and—very importantly—corporate trainers. (However, some newer products that attempt to radically simplify both deployment and usage need far less IT oversight).

Open-Source Full-Stack Products

The primary reason for choosing open-source BI tools is often perceived cost. Commercial BI tools are still largely seen as having superior technology, while open-source tools are viewed as offering good-enough technology at a fraction of the price. But although download of the software can be completely free, large-scale open-source deployments can still turn out to be a significant investment when factoring development costs. Also, there are very often commercial versions of the products that offer capabilities that the core free product does not. These typically include enterprise-level features like integrated security, connectivity to multiple data sources, administration tools, etc.

It is also important to bear in mind that these are developer-led tools and are designed with a developer mindset, which often means that significant development resources will be required to deploy and integrate them in an existing corporate environment.

There is however renewed interest in open-source BI tools today, partly fuelled by the extraordinary success of open-source products like [Hadoop](#) and [Revolution Analytics R](#), recently acquired by Microsoft, which have raised awareness of the open-source approach.

Example Products

- » [Actuate](#) (commercial product built on open-source technology)
- » [Pentaho](#)
- » [Jaspersoft](#) (acquired by TIBCO in 2014)

Best Fit For

- » Open-Source BI can be a good choice for organizations that have the technical expertise required to integrate the code base and make it work effectively within the organization. Typically these tool sets are very complete, due to the large number of developers working on the code base.
- » Open-source reporting engines are a particularly logical case for application vendors looking for a reporting engine to integrate into their product.

Cloud Full-Stack Products

Cloud full-stack BI products are a subset of full-stack products. They tend to include a data store, an ETL and semantic layer, and a range of front-end presentation tools sitting on top. The difference is largely in the deployment model (cloud versus on-premise). However, it makes sense to consider these solutions independently since they have some unique characteristics. For example, they are far easier to deploy, and do not require nearly as much IT oversight as traditional full-stack BI products.

Increasingly traditional full-stack BI providers are offering cloud versions, but most are single tenant, i.e. a single instance of the software supporting a single customer. Cloud-only full-stack BI products like Birst and GoodData are true multi-tenant SaaS products deployed on public clouds and offer all the advantages of true SaaS products, i.e. lower cost, frequent updates, and no data center infrastructure required. Tableau introduced a cloud-version of the product, Tableau Online, in 2013.

This familiarity with the cloud paradigm for enterprise business systems, in conjunction with massive and growing demand for analytics by business users, has spurred the development of cloud BI systems, and a basic comfort level with analytics in the cloud. It was inevitable that as operational data moved to the cloud, analytics would soon follow, and cloud BI is fast becoming ubiquitous, despite some reservations among highly regulated industries. Cloud BI platforms are often positioned as splitting the difference between legacy and discovery tools, offering the ease-of-use of discovery tools, with legacy data integration capabilities.

Cloud BI has been talked up as the next big thing in the BI world for some time now, but adoption has been slower than expected. One of the major obstacles has been concern over data security—corporations have been reluctant to put sensitive data in the cloud. However, as

more operational data is located in the cloud as cloud-based operational systems like Salesforce, NetSuite, Zendesk, SuccessFactors and a multitude of others become ubiquitous, cloud BI adoption is becoming much more mainstream. This familiarity with the cloud paradigm for enterprise business systems, in conjunction with massive and growing demand for analytics by business users, has provided a basic comfort level with analytics in the cloud. It was inevitable that as operational data moved to the cloud, analytics would soon follow, and cloud BI is fast becoming ubiquitous, despite some reservations among companies in highly regulated industries.

Example Products

- » [GoodData](#)
- » [Birst](#)
- » [Domo](#)
- » MicroStrategy Secure Cloud

Best Fit For

- » Organizations that have come from the Internet world and have been using SaaS applications like Salesforce and SuccessFactors to run their businesses. These organizations are likely to have fewer security concerns around storing their data in the cloud.
- » Organizations of all sizes that want a much easier to deploy, less IT-centric version of the full-stack products allowing “single version of the truth” reporting across a department or a whole company.
- » Smaller organizations with a limited budget that want a fully featured system at far lower initial cost due to the absence of any capital outlay for on-premises infrastructure.

Data Discovery and Visualization Products

Data discovery and visualization tools are designed for data analysts and more technical business users. The focus of these tools is not really reporting and monitoring, but rather ad-hoc analysis of multiple data sources. They provide data analysts with an intuitive way to sift through large volumes of disparate data to expose patterns and outliers hidden in the data. They replace the traditional rows and columns of traditional data presentations with graphical pictures and charts.

These tools have taken the BI world by storm, largely because of the low cost of implementation and because they do not require IT support. Ease of use is another key feature encouraging rapid adoption. They allow end users with some comfort level in data analysis to access multiple different data sources and perform data mash-ups and display the results in visually compelling ways. For example, a company might produce a visualization of expenses by department across a large enterprise to help hone in on outliers and figure out the reason for the disparity.

Example Products

- » [Entrinsik](#)
- » [IBM Watson Analytics](#)
- » [QlikView](#) and [Qlik Sense](#)
- » [MicroStrategy Analytics Desktop](#)
- » [SAP BusinessObjects Lumira](#)
- » [Tableau Online, Desktop & Server](#)
- » [TIBCO Spotfire](#)

Best Fit For

- » Business analysts requiring access to data from disparate systems, and the ability to blend the data with no required IT assistance, and produce visually compelling images to understand the data and tell a story.
- » These are not the right tools for providing a reporting infrastructure across an entire company and very few companies use these tools as their corporate BI standard, but it is also rare for at least one of these tools not be used at an individual or departmental level.

Big Data Products

Big data does not describe a single technology or initiative, but rather a broad trend that is affecting all kinds of organizations. Big data technology emerged in response to the enormous volumes of data that have inundated organizations in recent years, and that are beyond the capacity of traditional business intelligence tools to process and manage. The problem that Big Data technology vendors are trying to solve is how to actually use this data to improve business outcomes. Terabytes of digital information are collected from actual physical devices like RFID sensors and machines, along with human-sourced communications like text image or video. Most existing BI systems cannot easily comprehend this kind of data, as they have been designed to make sense of highly structured data organized in tables and stored in a data warehouse. That leaves a vast quantity of potentially very useful data out in the cold. This is the driver behind the rapid ascension of the Hadoop and noSQL data stores like [MongoDB](#) and [Cassandra](#), and the constellation of products that have accrued around them.

The value the big data technology can bring to the enterprise is varied and profound. Here are some typical use cases among many:

- » **IT Data Center Optimization:** Running a large, complex modern data center is not an easy task. A large data center can produce terabytes of plain text log files. Big data systems can help analyze this massive volume of log files to understand the root cause for any system

breakdown, or sub-optimal performance. These systems can analyze terabytes of data daily to decipher what is happening across the stack with every single transaction. Without big data systems, this is impossible.

- » **Fraud Detection:** Fraud detection is all about building models in order to identify customers engaging in fraudulent behaviors. The problem is building these models however, is the underlying data. Because the volume of transactional data is so intimidatingly large, models are usually constructed on subsets or segments of the entire data set. Partial data and high latency can seriously reduce the predictive power of these models. Big data tools allow models to be built on the entire data set and with very low latency, thus vastly improving the power and accuracy of the predictive models.
- » **Call center analytics:** Big data models can help to understand customer loyalty decay, and to remediate customer dissatisfaction at key touch points to increase customer loyalty.
- » **Social media analytics:** Analysis of torrents of data in the form of social media streams can provide insight into what customers are saying about a company and its products along with those of competitors. While this sentiment analysis is important, the real power of social analytics is linking this sentiment data to transactional data to understand how sales promotions, loyalty programs and competitor activities correlate to this social sentiment.

Example Products

Hadoop Infrastructure

- » [Hadoop](#)
- » [Hortonworks](#)
- » [MapR](#)
- » [Amazon Elastic MapReduce](#)
- » [Microsoft Azure HD Insight](#)

Big Data Analytics

- » [Platfora](#)
- » [Datameer](#)

SQL on Hadoop

- » [Apache Hive](#)
- » [Apache Spark](#)

Best Fit For

- » Companies who need to analyze very high volumes of data, from very diverse data sources, to solve pressing and complex business problems.

- » Data-rich organizations with an IT department closely connected to business units, and with a strong desire to use their data to gain competitive advantage.

EMBEDDED BI

Several BI vendors also sell their products to ISVs to embed analytics capabilities in their own products. While this is not a separate class of products, this specific use case has become increasingly important as companies grapple with ever-growing data volumes and become more familiar with data discovery and visualization tools. Many software vendors realize that built-in analytics capabilities are critically important to the success of their products in the marketplace and are faced with a critical build-or-buy decision. They can add significant competitive advantage to their own customers by providing tested BI technology. This can allow them a rapid time to market and better cost management than building capabilities in a domain where they have limited expertise.

In the early days of this model, companies embedded proprietary code in their products using APIs provided by BI vendors. However, web-based solutions no longer need to be embedded, but reside adjacent to the application, greatly simplifying deployment and administration.

The market initially developed around open-source products designed primarily for developers like Logi Analytics, Tibco Jaspersoft, OpenText and Pentaho (Hitachi Data Systems). The emergence of cloud hosting infrastructure allows vendors to make their solutions available in platform-as-a-service cloud environments.

This use case is quickly becoming pervasive in the marketplace. GoodData, for example, is now almost exclusively focused on this market. In addition to GoodData, embedded solutions are also available from Logi Analytics, Tibco Jaspersoft, OpenText and Pentaho (Hitachi Data Systems), Birst, Qlik, Tableau, Looker and Sisense.

Recent Trends in the Business Intelligence Market

Shift from Centrally Governed to Agile, Self-Service BI

It's impossible to discount the extraordinary success of data discovery and visualization tools, and the emancipation of business analysts from old-school ETL processes and data modeling. This revolution has democratized data and greatly accelerated the speed of data analysis to help companies make data-driven decisions in fast-moving, highly competitive environments. The structured world of data warehouses and ETL processes as the single source of truth within the enterprise has been permanently challenged.

The hard distinctions between the various types of BI tools described in the section above are becoming less significant, as all vendors orient their products away from the IT-centric model towards a more agile, self-service approach. For example, full-stack vendors have all now built data discovery and visualization tools as a component of their platforms, and these new capabilities are improving all the time:

- » SAP Lumira started life as SAP Visual Intelligence and was renamed SAP Lumira in 2013, and SAP BusinessObjects Lumira more recently. The product has gone through several enhancement cycles and has become a solid discovery and visualization tool for SAP Business Objects users who use the two products together. Additionally, SAP has built an entirely new cloud-based BI platform called Cloud for Analytics, incorporating not just standard BI components, but also business planning, predictive analytics, and [Governance, Risk, and Compliance](#) (GRC).
- » MicroStrategy 10, and subsequent point releases, has also significantly improved the discovery and visualization tool called Analytics Desktop, since its initial release in 2013. Since then the company has updated the licensing model, usability and governed data discovery capabilities.
- » IBM released Watson Analytics at the end of 2014. This is an entirely new cloud-based analytics platform designed to process natural language queries, pattern detection, and data discovery with advanced analytics. The product has already amassed a significant user base through the freemium sales model and is starting to acquire paying customers. IBM is now applying Watson user experience and design principles to the Cognos platform, and the latest release of Cognos has been given a new name: Cognos Analytics.
- » SAS released Visual Analytics in 2012, and this has now become the flagship product. The Enterprise BI Server product still available for large deployments but is no longer a focus for the company.

- » Microsoft released the second major version of Power BI in 2015, and continues to enhance its capabilities. The product is a cloud-based data discovery and visualization framework with both desktop and browser-based authoring. It has pre-built connectors to 60+ different data sources.

Conversely, data discovery and visualization vendors are being pushed in the opposite direction. Products like QlikView/Qlik Sense, Tableau and TIBCO Spotfire comprised the first entirely successful attempt to wrest analytics away from the control of the IT department and make these capabilities available to business users who no longer have to rely on the IT department for data analysis. The IT department typically has sophisticated data management expertise, and for this reason has functioned as the de-facto BI service bureau for business departments that do not necessarily have those skills. But as business has speeded up and the urgency of understanding data has increased, the IT department became a bottleneck hindering the ability of business units to make data-driven decisions.

Ironically, the very success of this data discovery and visualization movement has sewn the seeds of reaction. A common scenario is that multiple groups within an organization purchase more and more seats for a data discovery and visualization tool, which began as a small departmental purchase. This inevitably leads to calls for enterprise licenses, and once a product becomes a standard offering deployed across the enterprise, the IT organization is inevitably involved again. Data discovery and visualization tool vendors have responded by building enterprise features like security, data governance, data preparation, and even report generation into newer versions of their products in order to satisfy the requirements of the IT department.

“ Qlik and Tableau started by selling into the business side directly, many times by-passing IT. It became clear that multiple instances could cause fractured environments with inconsistent results produced. Each department could develop its own numbers which often did not mesh. There was no single source of data. Eventually customers wanted enterprise licenses, which means the vendor had to deal with IT. The IT department gets involved and brings in questions about governance, metadata, administrative functions. That’s when Qlik and Tableau had to add more sophisticated data management functions and features. Qlik bought Expressor and integrated it into their Sense product for easier data integration / metadata management and Tableau now has its own data preparation and metadata management capabilities. Both companies have enhanced their governance and audit capabilities as well. They are now able to satisfy most of the IT requirements for enterprise implementations.



Claudia Imhoff
CEO Intelligent Solutions, Author



Tableau and Qlik are now being forced down the same road as traditional vendors, with increasing demands for enterprise level data governance, and even report generation. Ultimately, the question of whether the data is clean and reliable becomes central.

Barry Devlin

BI Consultant, Author, Speaker



Increased importance of Big Data Systems

The ascendance of big data and Hadoop is another major thread in the development of the business Intelligence landscape.

Hadoop Adoption Rate

2015 saw increased adoption of Hadoop and Hadoop-related tools. This is not exactly a new technology: Hadoop has been around for 10 years, and is still only being used by a relatively small number of early adopters. However, a Hadoop adoption **survey** based on 2,200 responses conducted by AtScale indicates that of those who already use Hadoop, 76% plan on doing more within the next 3 months. Of those who have not yet deployed Hadoop, almost half say that they plan to do so within the next 12 months. In addition, 94% of respondents are bullish about their ability to achieve value from Hadoop. Hadoop does look as if it's poised for significant growth and adoption.

How Does Hadoop Differ from a Traditional Data Warehouse?

Hadoop is often erroneously thought of as a database. It is, in fact, an ecosystem of open-source components including MapReduce, the Hadoop Distributed File System (HDFS), the HBase NoSQL database, along with other databases, and many other packages facilitating import and export of data into and out of the HDFS. All of this software is deployed and run on inexpensive commodity hardware—usually many different servers—to cope with the massive volumes of data.

One of the most significant differences between Hadoop and a relational data warehouse is the way in which the data is stored. In a data warehouse, the data is carefully structured and organized before it is stored, so the data is highly structured and easily accessible through well-constructed queries. A Hadoop data lake by contrast contains large volumes of raw, unstructured data, which can be analyzed by business analysts and data scientists without the constraints of any preconceived structure being imposed on the data.

The terminology often used to describe this difference is “Schema-on-Write” versus “Schema-on-Read”. In Schema-on-Write, the data is mapped and parsed before being written into predefined columns and rows in the warehouse. Conversely, in Schema-on-Read, analysts can use tools like Hive, Spark and other similar tools to analyze the data in its native format. Another way of putting this is that ETL is performed on the fly. There are advantages and disadvantages to both approaches, but one of the big advantages of Schema-on-Read is the ability to analyze raw, unstructured data without being slowed down by an existing structure or schema that may inhibit creativity and flexibility.

Why Hadoop Matters

In the first edition of this guide, we described the problem that big data vendors are trying to solve: How to harness the Terabytes of unstructured data like streaming data, video data, machine data, etc. to improve business decision-making and business outcomes. Business data is no longer only collected from operational and transactional, internal systems, but also from physical devices like sensors and machines, and from human sources like social media, image designers, etc. The relational data warehouse was designed for highly structured data stored in tables, and cannot comprehend this kind of unstructured data, or this volume of data—hence, the rapid ascension Hadoop and of so-called “data lakes,” or vast repositories of raw data stored in its native format until needed.

BI tools must all now be capable of ingesting and analyzing this data, often in conjunction with more organized, structured data. Virtually all BI vendors now integrate with Hadoop in some fashion, and many legacy BI vendors have formed partnerships or acquired vendors in the big data space to be able to tap into this new data universe. Notable big data acquisitions include:

- » Teradata acquired four Hadoop-related companies in 2014: [Think Big](#), [RainStor](#), [Hadapt](#), and [Revelytix](#)
- » IBM acquired two healthcare big data companies, [Explorys](#) and [Phytel](#), in 2015 as it builds out healthcare big data analytics capabilities on Watson. It also acquired [Cleversafe](#), a big data storage product.

Microsoft acquired [Metanutix](#), a startup designed to help people crunch big data, in December of 2015.

More deals of this kind are likely in the second half of 2016 and in 2017.

Data Preparation and Machine Learning

The two trends just described: a shift away from IT-managed BI deployments towards agile, data discovery and visualization tools, and an increasing emphasis on schema-less Hadoop data lakes, have both led to a third major trend: data preparation and machine learning.

The task of data preparation used to be performed by IT departments running Extract, Transform and Load (ETL) processes. ETL extracts the data from the various data sources, and loads it into a data warehouse where it is normalized by organizing it into tables, while cleaning the data and removing redundancy and inconsistencies. Once it has been appropriately structured, it is then available for querying and analysis. In the new self-service, agile world, this paradigm no longer holds. As data becomes more democratized, one of the biggest challenges for business users trying to make sense of data for analysis, is that the data must first be prepared. Data from multiple different sources has to be integrated and cleaned before any analysis can occur. How successful less technical business users (rather than data analysts and data scientists) can be at this task is a matter of debate. But given the scarcity of technical data scientists and analysts, the goal of many vendors is to create a kind of “ETL light” that requires the minimum amount of expertise in order to be successful.

Data Preparation for Data Discovery and Visualization Products

Data discovery and visualization tools like Tableau and Qlik have typically relied on third-party tools like Alteryx to clean and prepare data for analysis. Indeed these product types are quite complimentary, and the vendors have lead sharing agreements in place. However, data discovery software vendors are increasingly developing their own data integration and data cleaning capabilities. For example, Tableau 9 introduced some Excel-based data preparation capabilities that are a first step in that direction. Qlik has always had the ability to perform data loads and basic data preparation tasks though scripting, but non-technical users have been forced to rely on third-party tools. TIBCO Spotfire 7.6 has what it calls visual “in-line data wrangling” functions that lets users perform data preparation functions while performing their analysis, an approach they believe is more useful than workflow style preparation tools. Ultimately though, all of these vendors are likely to build robust data preparation capabilities into their products so that users are not forced to purchase separate products to perform this mechanical but crucial process.

But the jury is still out on the ability of business users to use these tools successfully.



Tableau and Qlik are terrific additions to the marketplace; they expanded the ability of people to be able to analyze data. But while these tools have made a great leap forward, I don't see a comparable leap on the data side – if you look at wrangling tools and data prep tools, they are still aimed at techies, or IT people, or consultants like me. We're not really quite there yet.



Rick Sherman
Managing Partner at Athena IT Solutions

A more likely scenario is that data analysts or even data scientists will still be the primary users of these tools, at least in the medium term. The increasing prevalence of machine learning technology may eventually bridge this gap, but this is still an issue in the current environment.

Data Preparation for Big Data

Big data represents an even bigger challenge. The explosion of interest in big data technology, as organizations begin to understand the potential competitive advantage to be gained from analyzing massive quantities of unstructured data, has triggered a burst of innovation across the entire analytics landscape. As pointed out, data warehouses and big data stores like Hadoop are vastly different, but the need to prepare the data for analysis is equally crucial for both scenarios. This is and has always been an arduous task and frequently takes longer than the time required to actually analyze the data.¹

One direct result of this is the emergence of a new class of data preparation tools like the aforementioned Alteryx, specifically designed for big data preparation: [Trifacta](#), [Paxata](#), and [Tamr](#), are some of the newer entrants in addition to more established vendors like [Informatica](#), [Datawatch](#), and [IBM](#).

¹ For Data Scientists, “Janitor Work” is Key Hurdle

When IT was the steward of enterprise data, specialists handled data preparation and integration as part of the ETL process. The difference today is that data preparation tools are no longer being designed for the IT specialist, but rather for the self-service data analyst or business user.

Machine Learning

Machine learning is a relatively new data analysis method that has become a hot topic in analytics generally, but is also getting a lot of attention in the context of data preparation. As data analysts and even business users with limited data management expertise are now frequently performing data preparation on the fly, it becomes critically important to build software that can intuit and understand large volumes of data automatically. Machine learning technology uses algorithms that are capable of learning iteratively. Given that business users and data analysts now have to perform their own data preparation without assistance from IT, modern data preparation tools have started to build machine learning under the hood, in order to make things as easy as possible for non-experts to perform things like data integration and format conversions on their own. For example, the software can suggest tactics to users for blending data or other data preparation scenarios based on what others have done before. By drawing on a library of past actions, the software is capable of guiding the user to accomplishing tasks that might otherwise have been too complex or require the assistance of a data expert.

Coexistence of Multiple Data Sources

Many large enterprises are making big bets on Hadoop as a critically important data framework for the future. This technology will be indispensable for managing huge volumes of unstructured and semi-structured data in a cost-effective and highly flexible environment. However, it should not be understood as a replacement for the traditional data warehouse. This is not a rip-and-replace technology. According to data warehouse pioneers like Barry Devlin and Ralph Kimball, the two technologies will exist side-by-side for the foreseeable future. Hadoop will be used for mass storage of unstructured data for predictive and exploratory data discovery. Data warehouses, particularly those running against very fast, massively parallel processing (MPP) relational databases like Redshift, Vertica, Netezza, and Google BigQuery, will remain the best infrastructure for structured reporting, which is still the lifeblood of most organizations. As mentioned above, a new breed of data warehouse automation tools have mitigated some of the drawbacks of setting up and managing these data stores.

Ralph Kimball described this coexistence succinctly in a webinar in 2014: “Everyone has now realized that there’s a huge legacy value in relational databases for the purposes they are used for. Not only transaction processing, but for all the very focused, index-oriented queries on that kind of data, and that will continue in a very robust way forever. Hadoop, therefore, will present this alternative kind of environment for different types of analysis for different kinds of data, and the two of them will coexist.”²

² *Building a Hadoop Data Warehouse, Dr. Ralph Kimball, 2014*



The data warehouse is where you create production analytics. These use trusted, high quality data, the data you count on for KPIs, regulatory and compliance reports, precise financial analyses, and so on. But there is also a need for an experimental environment. This environment uses (big) data without formally vetting it, without rigorous data quality processing or even data integration processing. Data analysts and data scientists just want to experiment with the data, try different analytical techniques, perform general, unplanned queries and analyses. This environment is what I call the Investigative Computing Platform. It is not as rigorously controlled as the data warehouse and has more flexible governance and schema support. These two environments (the production data warehouse and the more experimental investigative platform) are diametrically opposed to each other but serve important purposes in the world of analytics. Will there be a single technology that can handle them both? The technologies are certainly moving in that direction, but perhaps not today; I believe it will be a while before we see the data warehouse and the experimental environment fully supported in a single technological environment.



Claudia Imhoff
CEO Intelligent Solutions, Author









The data infrastructure of the future will include a variety of data repositories including relational database servers, Apache Hadoop, and other NoSQL platforms, interlinked by a metadata catalog defining the characteristics and context of all the data in each store.

Data warehouses and standard BI reporting are not going to disappear any time soon.










Implications of Trends for BI Buyers

- » Traditional, data warehouse-based BI systems sometimes referred to as “legacy BI” are unlikely to be at the top of the modern BI buyer’s list, at least to begin with. Although these systems are remarkably effective within their domain, they have traditionally not been able to supply the agility and exploratory freedom that is required by today’s business environment. However, many of these vendors are remaking their product suites to provide more agile capabilities and reduce dependence on the IT department. These products are likely to remain good options for larger enterprises, as they continue to re-design their products to meet the needs of the modern agile enterprise.
- » Data discovery and visualization systems provide the agility and freedom missing from traditional tools, but have their own shortcomings. These very powerful tools tend to lack data preparation and data governance capabilities which are required if they are to be used as the primary BI system across an enterprise. The addition of a data preparation tool will almost certainly be needed to blend and structure data from multiple sources, although eventually these capabilities are likely to be built-in.
- » It is almost certainly unwise to swap out data warehouses and their associated BI systems for the new world of Hadoop and data lakes. The big data Hadoop infrastructure is designed to house high-velocity and high-volume unstructured data generated by machines. However, Hadoop makes little sense—at least today—as a repository for business-critical, highly structured, core business data, which is best, stored in a data warehouse or other structured data store.






BI Products Featured in this Guide

PRODUCT NAME	tr Score	BEST FIT FOR...
 ★★★★★ (31)	8.7 out of 10	Alteryx is a good option for companies with a need for a data blending and data modeling before staging the data for visualization in another tool like Tableau, QlikView, Microsoft Power BI, or TIBCO Spotfire.
 ★★★★★ (83)	7.7 out of 10	Birst is a pioneer in multi-tenant, cloud BI and a good choice for organizations looking for a cloud-based platform for business intelligence and analytics, that can help organizations looking to analyze complete business process, and operationalize analytics across both decentralized and centralized teams.
 ★★★★★ (30)	7.5 out of 10	Chartio is designed for business users and allows visualizations to be built by direct connecting to databases. The primary target audience is smaller companies.
 ★★★★★ (30)	8.5 out of 10	This is a relatively new product, and Dundas provides a lot of support for the embedded use case. Its broad range of functionality, and focus on a dedicated business user's experience make it a contender for companies looking for excellent visualization coupled with strong connectivity to multiple data sources and ETL.
 ★★★★★ (78)	8.5 out of 10	Informer is used by customers in a variety of industries where there is a strong need to manage large volumes of operational data. Typical customers have limited IT support and a relatively non-technical user base.
 ★★★★★ (29)	8.1 out of 10	GoodData's strength is as an embedded or white-labeled application under their "Powered By" program. This is targeted both to ISVs and to enterprises that wish to create and distribute analytics out to internal and external customers. This model makes up about 50% of the company's business, and is expected to grow.
 ★★★★★ (135)	6.8 out of 10	IBM Cognos Business Intelligence is an enterprise "full-stack" solution, and has recently undergone a major new re-design. The new platform is called Cognos Analytics. It is still the same platform under the hood, but the user experience is completely different and is similar to Watson Analytics. TrustRadius will endeavor to get early reviews of the new version to help provide guidance.
 ★★★★★ (34)	8.7 out of 10	InsightSquared is a good option for small to mid-size companies looking for a tool to support sales analytics, pipeline management, trend analysis, and sales forecasting, and for companies looking to more easily access and understand data from Salesforce.

Continued on next page.

PRODUCT NAME	trScore	BEST FIT FOR...
 ★★★★★ (28)	8.7 out of 10	Looker is a fast-growing new entrant and is a good choice for companies looking for a data exploration tool capable of ad-hoc data exploration of SQL-readable data storage systems. Originally more focused on tabular data, the company is rapidly developing front-end visualization features designed for business users. Looker is also frequently sold for its API, embedding, and OEM functionality.
 ★★★★★ (203)	7.6 out of 10	Microsoft BI is a good choice for large Microsoft shops, already using SQL Server across the organization. Power BI is a strong data discovery and visualization tool at a competitive price.
 ★★★★★ (99)	7.2 out of 10	MicroStrategy is a good option for medium to large enterprises in need of a complete set of business intelligence capabilities that can grow from a single data discoverer, to many thousands of users. It includes reporting, dashboards, and data discovery and visualization across the organization. It has particularly strong mobile capabilities.
 ★★★★★ (40)	7.5 out of 10	Pentaho is a solid open-source platform, and the data integration capabilities are particularly strong. This would be a good choice for companies that need to blend and integrate high-volume data from a wide range of structured and unstructured sources.
 ★★★★★ (230)	7.4 out of 10	QlikView is a good choice for companies with developer or power analyst resources to build reports and dashboards. Qlik Sense is a self-service tool built on the same engine that is being rapidly expanded as an enterprise platform.
 ★★★★★ (177)	7.0 out of 10	The SAP Business Objects platform is an enterprise-level system best suited to larger companies, and companies already using SAP enterprise applications. The new cloud platform, SAP Cloud for Analytics, will serve the same audience, although absence of deployment difficulties might extend its appeal to smaller organizations.
 ★★★★★ (64)	8.3 out of 10	SAP BusinessObjects Lumira is an effective visualization and data exploration tool for business users if they are already in the SAP ecosystem. It is easy to use, good value for money and provides excellent ROI.
 ★★★★★ (171)	7.6 out of 10	Crystal Reports is a good design tool for customers already using other SAP products, especially Business Objects, for which it serves as the de-facto report design tool. It is used in conjunction with SAP Crystal Server by mainly smaller companies for frequently repeated reporting tasks like quarterly sales data.
 ★★★★★ (35)	7.4 out of 10	Sisense is designed for companies with a need to process disparate data sources or large data volumes without heavy reliance on an IT organization. It also provides self-service analytics tools that are designed for the business user rather than the IT organization.

Continued on next page.

PRODUCT NAME	trScore	BEST FIT FOR...
 ★★★★★ (359)	8.1 out of 10	Tableau is an excellent tool for business analysts, allowing them to do discovery and visualization on data from a large number of data sources with ease. However, it requires additional products to help prepare the data for analysis. It is also not designed for enterprise reporting. TrustRadius does not have enough reviews to cover the Online version, but will endeavor to provide better coverage in the future.
 ★★★★★ (198)	8.0 out of 10	
 ★★★★★ (38)	7.2 out of 10	Jaspersoft is an open-source product suite, with particular strength in reporting and analysis, and an intuitive user interface. This is a good choice for companies looking for a reliable reporting engine that can be embedded into other applications.
 ★★★★★ (95)	8.7 out of 10	TIBCO Spotfire is a good choice for companies who need interactive visualization capabilities with built-in data wrangling, predictive and location analytics. Its cloud and on-premises platform is well suited for quickly building analytic applications that can be deployed across a small team or company.
 ★★★★★ (50)	7.8 out of 10	Zoho Reports is a suitable reporting tool for small and mid-sized companies looking for a relatively inexpensive SaaS BI solution with a simple UI and good connectivity to a wide range of databases and business applications.

Products not covered in this Guide

A number of BI products are not discussed in this guide because TrustRadius does not have adequate data. Some of the omissions include:

- » Actuate OpenText Analytics
- » Domo
- » IBM Watson Analytics
- » Information Builders
- » Microsoft Power BI
- » Oracle BI Foundation Suite
- » Qlik Sense
- » SAS Enterprise BI Server
- » SAS Visual Analytics
- » Targit
- » Yellowfin

How to Buy a BI Solution

There are a number of steps to completing a successful BI purchase and implementation:

Don't get hung-up on the technology too early. It's about understanding the business problem and having the right people and processes.

Buying a BI solution is a complex undertaking, not just because of the large number of options available, but also because there are many different kinds of products, designed to do very different things. But the technology is not even the most important thing. All of the experts we interviewed for this guide were unanimous: Don't start with the technology!

“ **Successful BI implementations are not just about tools, but are about people and processes. Focusing too soon on tool selection is not going to provide the best result. A BI project is not a one-shot thing, but is a journey that takes time and patience.** ”



Wayne Eckerson
BI consultant and Expert

“ **It's rarely about the technology. It's more around people and processes and scoping and managing projects properly. It's very important to start with the business and work backwards.** ”

John Onder
Principal CBIG Consulting



“ **The important thing to remember is that it's usually not about tools, and is more likely to be about organizational maturity and the ability to get decisions made and get things done. Many organizations I encounter are somewhat internally dysfunctional and this ability to get things done is weak or even non-existent.** ”



Barry Devlin
BI Consultant, Author, Speaker



It's about more than just the technology. You can implement a solution but if people can't take action or use the product to make progress in their job, then they don't use it and the project fails due to lack of adoption.

Lyndsay Wise

BI Industry Analyst, President Wise Analytics



If people and processes are more important than the technology, what kind of people and processes should be put in place? The following sections discuss the important of nominating an executive-level project sponsor and functional business leaders and the role of IT. However, there are some key success factors that are common to all successful BI projects. For example:

- » Create a business plan outlining the business problems to be solved and the expected benefits
- » Derived from the business plan, build a simple step action plan and outline each step clearly
- » Build a project team with all the appropriate stakeholders from business and IT
- » Establish clear success criteria

Find an executive sponsor

Perhaps the most important first step is to secure executive sponsorship with enough clout in the organization to telegraph the seriousness with which the organization is approaching the project. This is the most important overall factor.



Very often, organizations don't invest in the right people with the right knowledge and experience to know how to fit everything together. You need a very strong program manager, but also people who understand how to work cross functionally. Having the right sponsorship from the business side is essential.



Wayne Eckerson

BI consultant and Expert



To succeed, BI projects need a strong leader who is knowledgeable about both technology and business and can straddle both worlds, translating between the two. Since the ultimate goal is to achieve significant business value, it's usually better to have a technically-oriented business executive lead the team.

John Onder

Principal CBIG Consulting



Business should lead, IT should play a consultative role

Procuring a BI solution is a business decision. Only the business really understands the problems to be solved and the value that new technology can bring to the organization. Buying BI technology should never be an IT-only decision and this is something that needs to be made very clear from the start.

“ I draw a distinction between who should drive and who is usually driving when I get there! There is no question that a major project should always be driven by a high level executive on the business side; Either a CEO, or some direct report, who can take a cross-enterprise view. The reality is that IT is often reluctantly in the driving seat, and trying to get out of it as fast as they can. I try to transition leadership from IT and get the business people to step up to their responsibilities. IT is a co-owner in a process like this, but should not drive.



Barry Devlin
BI Consultant, Author, Speaker

“ BI projects should be driven by business. If they are driven by IT, there will be significant struggles. Speed of delivery is critically important. Often things will go on for far too long. A good idea is a hybrid agile approach where you mock things up and show the businesses how data can solve their problems... It's really very important to do this kind of rapid prototyping. The applications side of things is owned by business. The IT group is the curator and keeper of the data.

John Onder
Principal CBIG Consulting



“ Business users might want something not realizing that it will take six to nine months to implement rather than if they had chosen something else. IT usually prefers to choose something that fits into the already existing technology stack and tends to be less interested in business requirements.



Lyndsay Wise
BI Industry Analyst, President Wise Analytics

Focus on ease-of-use

BI tools can be notoriously difficult to use and it's important to understand the range of user types that will be using the software. The largest number of users are likely to be relatively non-technical executives, operations staff or salespeople who need the ability to monitor metrics, analyze anomalies and drill down to see details. A far smaller number will be technical users like data analysts or even highly-trained data scientists and modelers who really want to be able to explore

large data sets. It's important to understand the abilities of your users and to not overestimate the abilities of the largest part of the user population.

“[Organizations struggle and fail] because the technology is often hard to use, and they have not done proper due diligence around the products they have invested in. But also organizations struggle because they are so tied to the technology investments they have already made. They don't want to rip and replace what they already have. They are trying to upgrade what they are already using and get the performance that they need. Sometimes they need something newer which is more flexible.

Lyndsay Wise
BI Industry Analyst, President Wise Analytics



Start small and get a quick win

It's critically important not to try to boil the ocean, but tackle a small bounded product and show some quick results. For example if executives from different departments cannot agree on sales numbers because they have different definitions of "product" and are working from different spreadsheets, agree to track a couple of metrics to help solve that problem. Early success of this kind is the best way to build confidence in the program and ensure continued buy-in.

“ It's critical to get a quick win. Find a project of significant value to the business and deliver it quickly. Once the business gains confidence in the technical team, it will eagerly invest in additional projects. With momentum, the technical team can then lay the foundation for an enterprise-wide program.



Wayne Eckerson
BI consultant and Expert

“ Speed of delivery is critically important. Often things will go on for far too long. A good idea is a hybrid agile approach where you mock things up and show the businesses how data can solve their problems. Then they understand the concept of guided analytics. It's really very important to do this kind of rapid prototyping.

John Onder
Principal CBIG Consulting

