Accelerate BI Initiatives With Self-Service Data Discovery And Integration

Introduction

The rapid advancement of technology has ushered in a new era in which there are unprecedented amounts of data sources available to businesses — data that is ripe with insight if properly analyzed and understood. Businesses are keen on utilizing this data to better meet the needs of their customers, increase competitive differentiation, and improve business top and bottom lines. But reaping tangible benefits from business intelligence (BI) initiatives is easier said than done. With a highly competitive and ever-evolving market, BI practices must be agile enough to quickly respond to new business questions and needs. In the world permeated with data from Internet-connected devices and new avenues for data collection such as sensors and beacons, the vastness of the data available creates an overwhelming task of exploring and discovering the *right data* for every new BI initiative. Unless utilized fast enough, businesses risk the data becoming obsolete before they even have a chance to explore it for insights. Discovering the wrong or incomplete data can slow down the data exploration process and the BI initiative as a whole, which results in the inability to meet customer needs, loss of business opportunities, and diminished confidence in being able to use BI effectively.

In April 2015, Attivio commissioned Forrester Consulting to examine the challenges businesses face with BI initiatives and to assess the need for a more automated approach to finding the right data to be used in a BI initiative. This study looks at how companies are currently running BI projects and explores the needs and expectations for establishing more agile BI practices. To do so, Forrester leveraged data from its 2014 Business Technographics® Global Data And Analytics Survey and supplemented that data with a custom survey of 50 US-based IT and business decision-makers at large firms of 1,000 or more employees.

Business Intelligence Takes Too Long To Profile, Identify, and Unify Data Sources
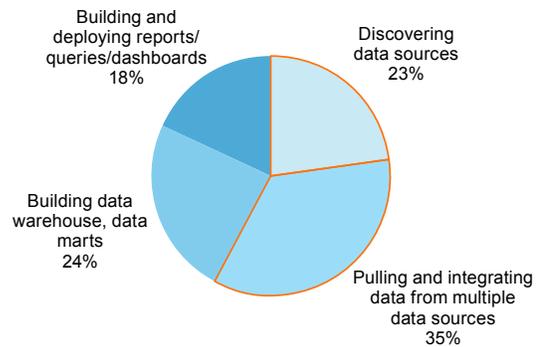
With the latest mobile devices and ubiquitous access to cloud-based products and services, customers are more empowered and connected than ever. This new environment generates vast amounts of data that, coupled with existing data sources, can provide tremendous opportunities for businesses to better understand their customers. However, despite the volume of data available, many businesses struggle with being able to quickly analyze data and turn it into actionable insight. According to Forrester research, 68% of simple BI content requests take weeks, months, or even longer to complete, and for more complex requests that number is 88% (see Figure 1). This is much too long by today's standard for any organization wanting to remain competitive and respond to increasing market demands, especially for companies that are seeking to establish more agile BI practices. The challenges of agility and flexibility present a significant hurdle in making BI truly effective. Customer needs and requirements are constantly changing, and unless businesses are able to quickly identify and respond to those changes, they will fall behind the competition.

So, why do current BI initiatives take so long to complete? What is preventing businesses from turning around BI requests much faster? In a 2015 Forrester report, Forrester explained that a good rule of thumb in any BI initiative is to assume that 80% of the effort is going to center on data integration activities. A similar 80% effort within data integration is spent just to identify and profile data sources,

thus leading to the assumption that *64% of any BI initiative is spent on identifying and profiling data sources*.[1] This estimate was supported by our custom survey findings that showed, on average, 58% of the effort for any given BI initiative was spent on discovering data sources and unifying data. Further, by including the steps to build an analytical data store, such as a data warehouse or a data mart, the number is pushed to an incredible 82% of the total project time and effort (see Figure 2). That's 82% of the total project effort before any actual analysis or exploration can even start.

FIGURE 2

Fifty-Eight Percent Of The Effort For BI Initiatives Is Spent On Data Exploration And Integration
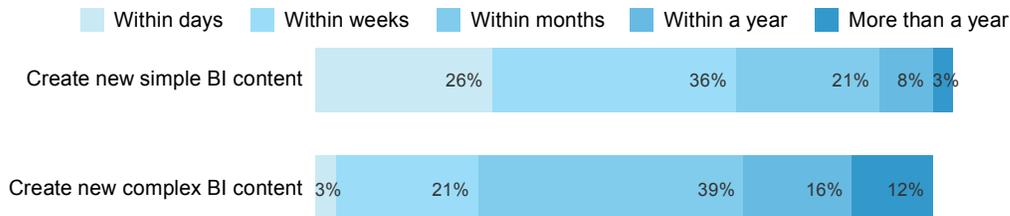

"How much of the total time spent to complete a traditional business intelligence initiative is spent on the following steps?"

Building and deploying reports/queries/dashboards 18%
Discovering data sources 23%
Building data warehouse, data marts 24%
Pulling and integrating data from multiple data sources 35%

Base: 50 IT and business decision-makers involved with BI and data management from US organizations with 1,000+ employees

Source: A commissioned study conducted by Forrester Consulting on behalf of Attivio, April 2015

FIGURE 1

Sixty-Eight Percent Of Simple BI Content Requests Take Weeks Or Longer To Complete

"In general, when business users are looking for help with BI, how quickly does IT turn around the following requests?"

	Within days	Within weeks	Within months	Within a year	More than a year
Create new simple BI content	26%	36%	21%	8%	3%
Create new complex BI content	3%	21%	39%	16%	12%

Base: 429 US technology and business decision-makers (1,000+ employees)

("never" and "don't know/does not apply" options not shown)

Source: Business Technographics Global Data And Analytics Survey, 2014, Forrester Research, Inc.

FORRESTER®

Improved Data Discovery Agility Is Needed

As the BI technology market has matured over time, there have been improvements in agility within various components of the stack, especially in self-service reporting, analysis, and data visualization technologies. But these improvements address the agility challenges of only 18% of any BI initiative. For most organizations, the data discovery component still remains a significant challenge. By bringing greater agility to the discovery and integration process, companies can reduce the time needed for BI initiatives, which would put them in a much better position to respond quickly to changing market demands.
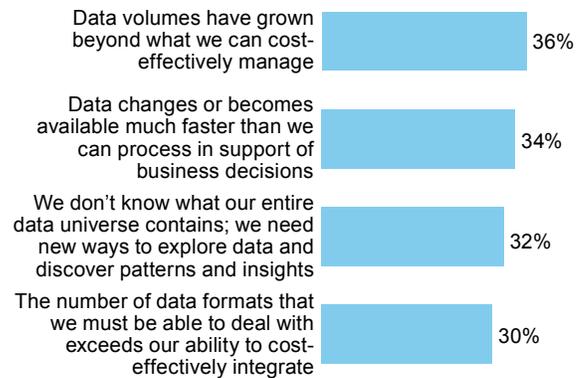
The extensive time it takes to identify, profile, and unify data sources can be attributed to the following data discovery challenges, each having an impact on nearly one-third of businesses surveyed (see Figure 3):

› **Data volumes are growing exponentially.** The sheer volume of available data tables, columns, and rows creates a daunting task of weeding through the clutter to get to the data sources needed for a particular BI request.

› **Data changes faster than businesses can process.** The newly arriving "big data" sources (i.e., sensors and beacons or mobile devices) all have the characteristic of changing dramatically from one day to the next based on customer behavior and market trends. This reinforces the need for agility in working with data in order to extract insights in a timely matter while they will still be relevant.

› **Businesses don't know what data they have and need new ways to explore it.** Given the challenges with data volumes and constantly changing data sets, BI users, analysts, and data scientists alike find themselves in a dilemma of not knowing what data they have or where to look for it. It is not practical to manually hunt through all the data available, so having a clear way to explore data more efficiently is critical to improving the turnaround time for BI requests.

› **An increasing number of data formats creates data siloes that businesses have difficulty integrating.** Having multiple data sources of multiple structures (structured, semi-structured, and unstructured) requires data to be stored in different types of databases. Without some sort of data hub or centralized data repository to integrate all of the enterprise data, data exploration will be limited to individual databases. The extra time needed to

FIGURE 3
Managing And Discovering Big Data Is A Challenge For One Of Every Three Businesses

"What are the main business and technical requirements or inadequacies of earlier-generation business intelligence technologies (e.g., relational databases and reporting tools) that lead you to consider new BI techniques and technologies?"
(Select all that apply)

Data volumes have grown beyond what we can cost-effectively manage	36%
Data changes or becomes available much faster than we can process in support of business decisions	34%
We don't know what our entire data universe contains; we need new ways to explore data and discover patterns and insights	32%
The number of data formats that we must be able to deal with exceeds our ability to cost-effectively integrate	30%

Base: 429 US technology and business decision-makers (1,000+ employees)
Source: Business Technographics Global Data And Analytics Survey, 2014, Forrester Research, Inc.

search through all the different databases is only part of a larger challenge, however. The much bigger danger of siloed data lies in missing valuable insights that would have otherwise been uncovered if the data stored within a nonintegrated database was brought into the exploration.

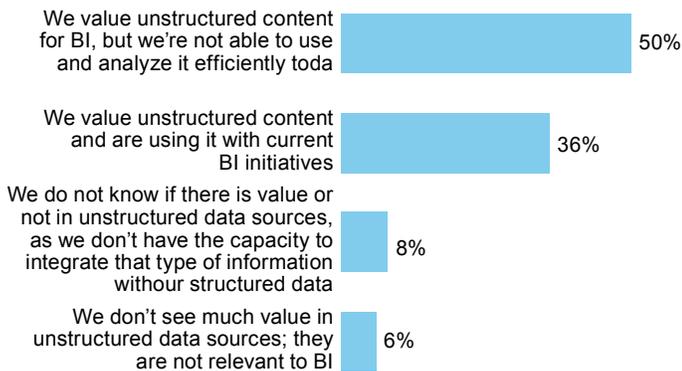## BUSINESSES WANT TO OBTAIN A FULL 360-DEGREE VIEW OF THE CUSTOMER

Forrester data shows that obtaining a 360-degree view of customers based on customer data is an important priority for 72% of businesses — most companies want to know everything they can about their customers in order to better serve them. While there is great advantage to having a holistic understanding of customers, it requires a herculean effort to integrate data from myriad sources. This challenge is exacerbated by the fact that an increasing number of the customer-centric data sources are in completely unstructured formats, such as call center logs, emails, and social media postings.

According to our custom survey, while an overwhelming 86% of businesses see value in using unstructured content for BI, only 36% are currently able to use unstructured

FORRESTER®

content effectively in their BI initiatives. Additionally, 8% of companies don't even know what value unstructured data could offer their business because they don't have the capacity to integrate it into their structured data repositories (see Figure 4).

"How do you see unstructured data fitting in with your business intelligence initiatives?"

We value unstructured content for BI, but we're not able to use and analyze it efficiently toda	50%
We value unstructured content and are using it with current BI initiatives	36%
We do not know if there is value or not in unstructured data sources, as we don't have the capacity to integrate that type of information withour structured data	8%
We don't see much value in unstructured data sources; they are not relevant to BI	6%

Base: 50 IT and business decision-makers involved with BI and data management from US organizations with 1,000+ employees
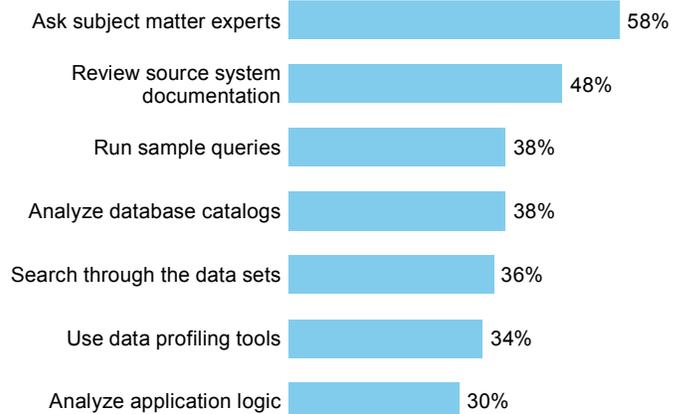
Source: A commissioned study conducted by Forrester Consulting on behalf of Attivio, April 2015

Self-Service Data Discovery Accelerators Are Critical For Meeting Agile BI Requirements

The desire to truly compete on analytics and insight generation requires more agile data exploration and data management capabilities to work with all data types, including unstructured data. One step to improving data discovery agility is to enable BI professionals with more automated tools and capabilities to explore and understand their data. Current discovery methods tend to be manual in nature and lack the automation and self-service capabilities needed to effectively profile large, expanding data sources. Our survey found that the top method for discovering new data sources was asking subject matter experts (58%), followed by reviewing source system documentation (48%) (see Figure 5). This dependency on others to identify the right data for context-specific analytics requirements causes huge delays in the discovery process, as it often requires

"How do you discover, understand, and evaluate data sources as inputs to your business intelligence applications?"

Ask subject matter experts	58%
Review source system documentation	48%
Run sample queries	38%
Analyze database catalogs	38%
Search through the data sets	36%
Use data profiling tools	34%
Analyze application logic	30%

Base: 50 IT and business decision-makers involved with BI and data management from US organizations with 1,000+ employees

Source: A commissioned study conducted by Forrester Consulting on behalf of Attivio, April 2015

multiple opinions and a trial-and-error path to ultimately find the best data fit.

Not all discovery and profiling processes are manual, however, as 34% of survey respondents indicated they were using data profiling tools. But these earlier-generation profiling tools have limitations, such as being:

› Dependent on trained technical resources

› Mainly focused on configuring data cleansing engines

› Heavily reliant on human-created metadata, which is often not updated as regularly as it should be and may not even exist in some legacy and newer NoSQL-based applications

As a result, these earlier-generation data profiling tools do not increase the productivity of the analysts performing data discovery exercises.[2] Forrester finds that BI data discovery agility can be enabled through a newer generation of "data discovery accelerators," which provide capabilities to:

› **Profile.** These accelerators automatically discover all enterprise data and content; analyze data and content to semantically enrich tables, columns, and files; classify data into categories such as person, customer, location,

FORRESTER®

and company; and evaluate quality metrics including cardinality, ranges, sparseness, null values, and outliers.

› **Identify.** These accelerators allow a searchable catalog of discovered and profiled sources; recommend new sources based on relationships with known sources; and use semantic search to handle ambiguity of terms and inconsistent data.

› **Unify.** These accelerators uncover relationships across data sets, spanning disparate databases and data and content types; autosuggest models based on referential integrity and join-ability from inferred relationships; and provision data to agile data discovery tools and advanced analytic engines.
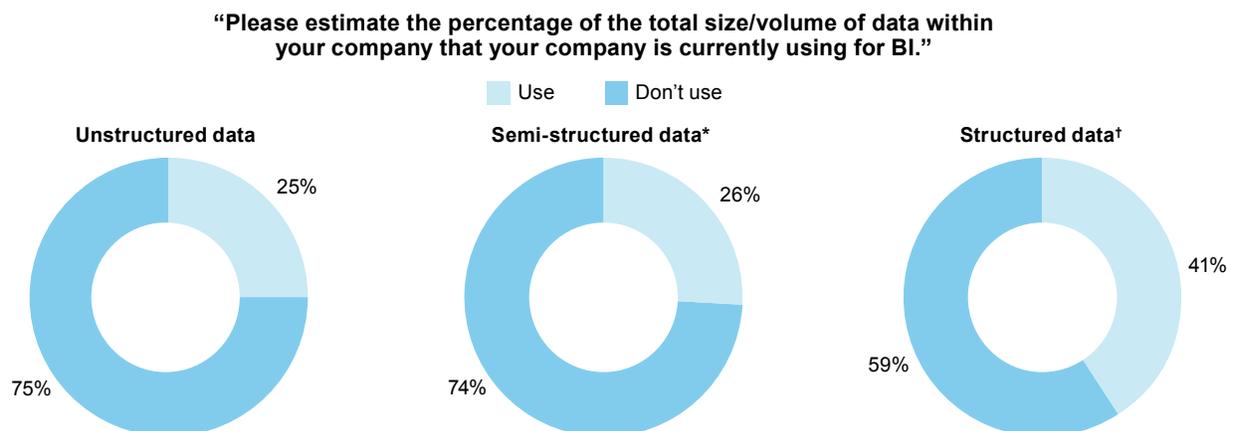
UNIFIED DATA ACCESS IS NECESSARY FOR AGILE BI

BI engagements can be completed faster and allow business to be more agile when they enable business analysts and users to better interact and explore the data more efficiently on their own, without the heavy dependency on data professionals such as data scientists, database administrators, or subject matter experts. Enabling this capability requires a universal but unified view of all the data sources available within the organization, which business users can explore and use on their own with little to no help from data professionals. For example, business analysts

usually can't script or parse files, and it would be impractical to open each file, one at a time, to find the right data source among thousands. Today these analysts are forced to rely on data management professionals to create custom integrated data views. This approach entirely takes away the promise of agile BI and BI self-service.[3]

Self-service data discovery accelerators provide enough visibility into all of your corporate data sources to allow for agile identification, unification, and access by the person who actually understands the context of how the data will be used to make a business decision. The added visibility into all available data sources can be a boon to BI initiatives, as Forrester data shows that businesses, on average, are only utilizing 41% of available structured data and just a quarter of data available in semi-structured and unstructured formats (see Figure 6). This demonstrates a clear need to better discover and unify data in order to ensure that more business opportunities or insights are uncovered. Data discovery accelerator tools will fill this gap, thus enabling true end-to-end agility in the BI process. These self-service data accelerators, which are specifically designed to scan through semi-structured Hadoop Distributed File System (HDFS) files, structured databases, and unstructured document stores, can automatically identify, profile, and metatag files and make them available for analysis and exploration by business analysts.

FIGURE 6
Businesses Are Utilizing Less Than Half Of Their Available Data For BI

"Please estimate the percentage of the total size/volume of data within your company that your company is currently using for BI."

Use   Don't use

Unstructured data: 25% Use, 75% Don't use

Semi-structured data*: 26% Use, 74% Don't use

Structured data†: 41% Use, 59% Don't use

Base: 250 technology decision-makers who know how much BI data their firm uses (1,000+ employees)

*Base: 229 technology decision-makers who know how much BI data their firm uses (1,000+ employees)

†Base: 255 technology decision-makers who know how much BI data their firm uses (1,000+ employees)

(These percentages are estimated based on ranges reported. These values are not exact.)
Source: Business Technographics Global Data And Analytics Survey, 2014, Forrester Research, Inc.

FORRESTER®

Conclusion

Achieving business success with big data and BI initiatives is a complex endeavor. Current BI initiatives take too long to build and deploy, and can result in failure to respond quickly to time-sensitive business or customer needs. Businesses know that big data and agile BI can offer tremendous value to their customers and business overall, but the steps of profiling and identifying the right existing data sources lack the much needed agility to discover and unify needed information efficiently. In response, business analysts must turn to new tools such as self-service data discovery accelerators to address challenges with the speed of data exploration and insight generation, the agility that new big data and unstructured data sources require, and the fear of missing an insight that unused data could have found. By adding self-service data source discovery and integration as the final piece to the puzzle, truly agile, end-to-end, self-service BI may finally be within reach.

Methodology

This Technology Adoption Profile was commissioned by Attivio. To create this profile, Forrester leveraged its Business Technographics Global Data and Analytics Survey, 2014. Forrester Consulting supplemented this data with custom survey questions asked of IT and business decision-makers at enterprises of 1,000 employees or more based in the US. Survey respondents were at the manager level and above. The auxiliary custom survey was conducted in April 2015. For more information on Forrester's data panel and Tech Industry Consulting services, visit www.forrester.com.

Endnotes

[1] Source: "Boost Your Business Insights By Converging Big Data And BI," Forrester Research, Inc., March 25, 2015.

[2] Source: "Continued Innovation In Business Intelligence: Data Discovery Accelerators," Forrester Research, Inc., June 21, 2007.

[3] Source: "Boost Your Business Insights By Converging Big Data And BI," Forrester Research, Inc., March 25, 2015.

ABOUT FORRESTER CONSULTING

Forrester Consulting provides independent and objective research-based consulting to help leaders succeed in their organizations. Ranging in scope from a short strategy session to custom projects, Forrester's Consulting services connect you directly with research analysts who apply expert insight to your specific business challenges. For more information, visit forrester.com/consulting.

FORRESTER®