A Forrester Consulting
Thought Leadership Paper
Commissioned By Pentaho

August 2015

Delivering Governed Data For Analytics At Scale

FORRESTER®

Table Of Contents

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®

Executive Summary

Data environments are growing exponentially as businesses evolve in an increasingly digital world. Structured and unstructured data accumulate from transactional, operational, content-centric systems and connected devices. Against this backdrop, businesses seek new and innovative ways to harness big data for decision-making, risk management, and business growth. Amid unprecedented data growth, how do businesses optimize their data environments for value? How do they ensure the delivery of trusted governed data as they integrate data from a variety of sources?

In June 2015, Pentaho commissioned Forrester Consulting to evaluate governed data delivery. To further explore this capability, Forrester developed a custom survey to test the assertion that organizations need to set up the right spectrum of data governance — i.e., a range of looser to stricter controls — based on the data's source, purpose, and amplification factor.

The right spectrum of data governance, along with supporting technologies and processes, is required to ensure data quality, accuracy, consistency, and ultimately usability.

In conducting an online survey of 164 business and IT professionals with data governance responsibilities at US and UK enterprises, Forrester found that organizations are scrambling to keep up with evolving business needs for data while trying to maintain quality and security. Highly effective firms take a more comprehensive approach to quality and struggle less with basic issues like completeness of data attributes and entities. As organizations' approaches to data governance mature, they can reap competitive benefits and improve data quality, in addition to managing risk.

KEY FINDINGS

Forrester's study yielded four key findings:

- **Enterprise-scale analytics capabilities require vast and varied data sources.** Companies are using a wide range of data sources, both internal and external, to power their analytics environments. On average, 52% of organizations blend more than 50 data sources to enable various analytics capabilities, and the vast majority use a combination of internal and external data. Our survey also revealed that levels of governance vary across data types. So when planning for analytic integrations, data professionals must consider quality and governance requirements across data sources and create approaches that ensure consistent, accurate, and trusted data insights.
- **Data quality is imperative and a major challenge.** More than 60% of survey respondents rated data quality, as well as security and privacy, as very important aspects of data governance. Quality and security are paramount but also challenging. Maintaining data quality and properly securing data topped the list of challenges firms face in governing data to help deliver reliable insights in their heterogeneous data environments.
- **Organizations that lack a strategic approach to quality fail to keep pace with business needs.** We compared Pacers, which we defined as respondents who indicated their organizations were effective and very effective at keeping pace with business data needs (59%), with the rest of respondents, which we categorized as Chasers (41%). We found that relative to Pacers, Chasers struggle to find and deliver data in a timely manner and face gaps in data completeness. Basic issues of filling data gaps impede productivity, making it difficult to improve capabilities. Organizations with a more strategic and comprehensive approach to governance and quality introduce flexibility to integrate and prepare data, enabling them to better meet a variety of business objectives (e.g., supporting a comprehensive view of the customer while ensuring risk and regulatory compliance).
- **Organizations that create a range of governance structures reap quality and competitive benefits.** We compared organizations that rated themselves high or very high in data governance maturity (64%) with those with lower maturity ratings. We found that mature organizations involve the business more in cross-functional governance models, while others lean on IT as de facto governance owners. Mature organizations assign ownership of governance activities to business, IT, or dedicated cross-functional governance teams based on where data comes from and how it is used. As a result, mature organizations use governance as a tool to drive strategic and growth imperatives, while others focus on more conservative benefits like compliance and risk reduction. Increased collaboration and business input direct integration and governance toward business capabilities that link data to specific tangible outcomes, transitioning organizations to a data-driven culture.

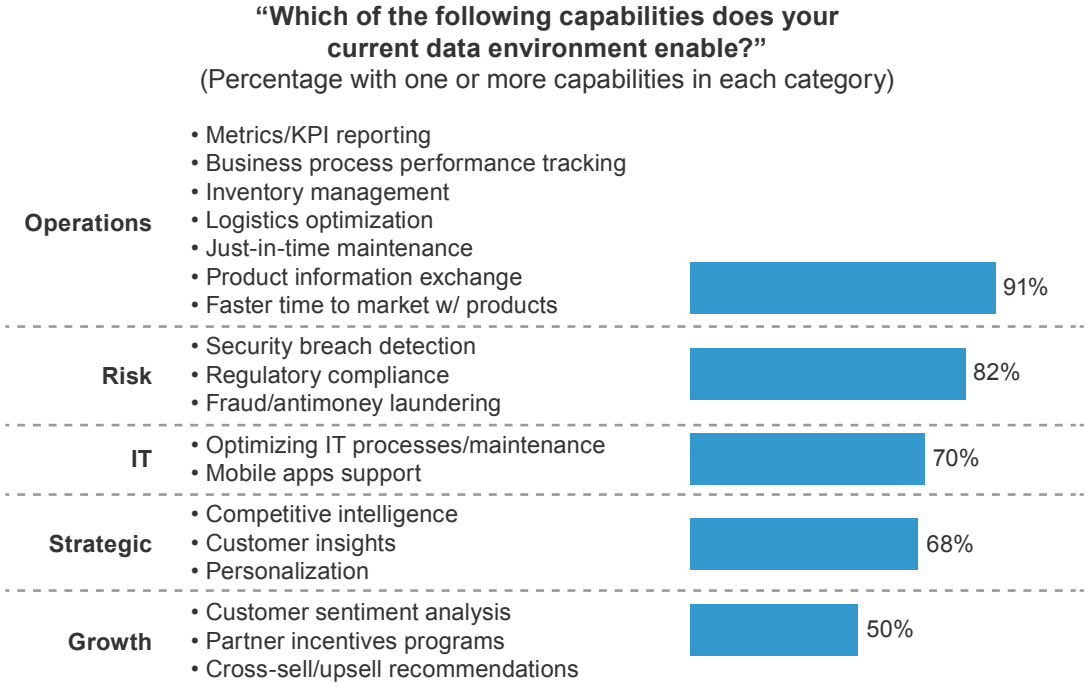
Vast And Varied Data Sources Power Today's Enterprise-Scale Analytics Capabilities

Businesses use data analytics to inform strategic decisions, power critical processes, communicate with investors and partners, and create exceptional customer experiences. Companies accumulate vast amounts of data every day from business processes and transactions as well as external sources to feed analytic efforts. In our survey of 164 data governance professionals, we found that transactional and unstructured data volumes exceed 100 terabytes (TB) each for the majority of US and UK enterprises. At more than 20% of companies, volumes exceed 1,000 TB each. How do companies put that data to use? Our survey revealed that:

› **Data governance professionals today prioritize operations and risk over reward.** The vast majority of data governance professionals have enabled one or more operations or risk capabilities — 91% and 82%, respectively. In contrast, 50% use analytics for business growth purposes such as customer sentiment analysis, partner incentive programs, or cross-sell/upsell recommendations (see Figure 1). While capabilities skew toward keeping the business operationally efficient and secure, as companies begin to expand their analytics capabilities to support business growth, they can capture upside benefits in addition to managing for downside risks.

› **Firms blend many data sources to generate analytics insights.** On average, 52% of firms blend together 50 or more distinct data sources to enable analytics capabilities. About a third (34%) blend 100 or more data sources, and 12% blend 1,000 or more (see Figure 2).[1] With the majority of businesses (53%) reporting in excess of 100 TB of unstructured data, many blended analytic environments likely include big data. When individual sources include automated and/or manual inputs, originate from disparate systems with different architectures, and are subject to different levels of governance, an effective integration process is essential.

FIGURE 1
Analytics Capabilities Skew Toward Operations And Risk Rather Than Growth

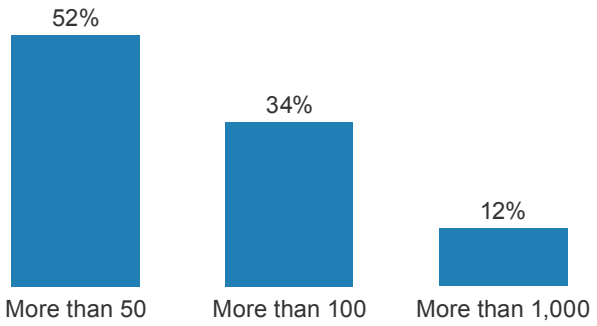


Base: 164 US and UK data governance decision-makers and influencers
Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

FIGURE 2

Most Firms Blend More Than 50 Data Sources In Their Analytics Environments

"How many data sources/feeds does your organization blend together to enable its enterprise-scale analytics capabilities?"



Base: 164 US and UK data governance decision-makers and influencers

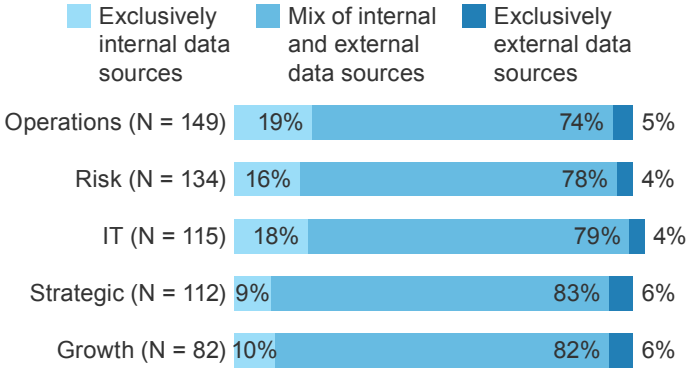
(Percentages shown represent averages across a list of capabilities. "Don't know" responses not shown.)

Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

FIGURE 3

The Vast Majority Of Firms Blend Internal And External Data Sources For Analytics

"For enterprise-scale analytics capabilities, what portion of the data comes from internal and/or external sources?"



Base: Variable US and UK data governance decision-makers and influencers

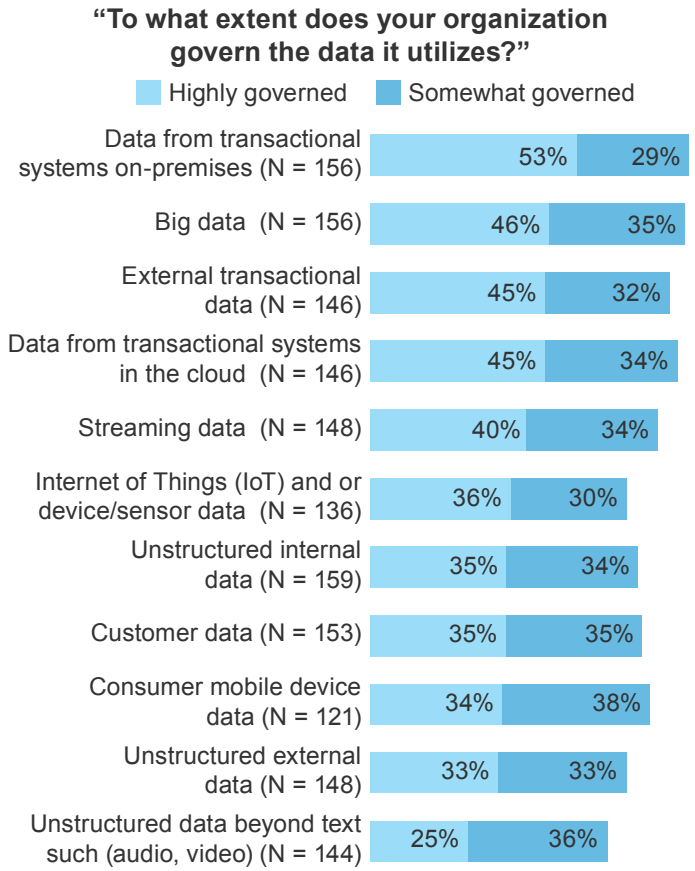
(Percentages shown represent averages across capabilities within each category. "Don't know" responses not shown.)

Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

› **Internal and external data power analytics capabilities.** We saw that organizations blend a high number of data sources, but where does the data come from? The vast majority of survey respondents indicated they use a mix of internal and external data sources across different types of analytics capabilities (see Figure 3). External data sources can introduce high volumes of unstructured and semistructured data. For example, we asked survey respondents the extent to which their organizations utilize various data types, and found that the majority at least moderately utilize unstructured external data from sources like social media (66%), Internet of Things or device/sensor data (65%), or consumer mobile device data such as geolocation and wearables (58%). This adds another layer of complexity and begs for data sources that are clean and complete. Data sources must be governed individually as well as in combination with other sources during integration to populate the authoritative hub for analytics.[2]

› **Different types of data require different levels of governance.** Data professionals recognize that all data is not created equal. As organizations aggregate various data sources, they are inherently blending data with different originating data governance standards. As expected, on-premises transactional data receives the most governance, with 53% highly governed, as it is highly utilized and operationally critical. Big data, external transactions, and cloud sources are also highly governed by 45% to 46% of organizations. However, customer data that contributes to an organization's strategic and growth objectives is still in a relatively chaotic state, with only 35% high governance (see Figure 4). Forrester's research indicates that customer experience takes top priority.[3] Thus, data professionals will need to invest more effort in governing customer data to glean reliable insights that drive topline business objectives. They must consider upfront which governance structures to apply. This ensures flexible integration and preparation of data, allowing organizations to apply quality standards as and when appropriate — even if the answer is that data will remain in raw format, waiting for governance when data is requested.

FIGURE 4
Governance Varies Widely By Data Type



Base: Variable US and UK data governance decision-makers and influencers

Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

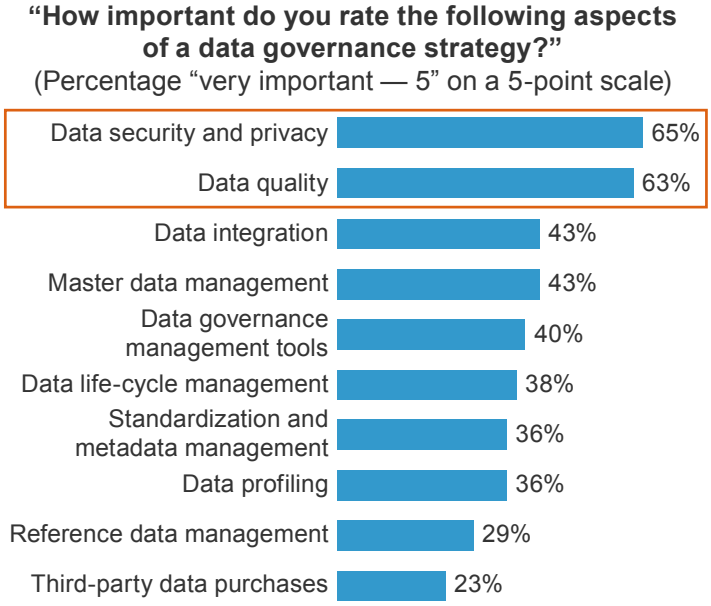
Data Quality Is Imperative But Also A Major Challenge

Data environments power organizations' core business processes, manage risk, inform strategic decisions, and enable growth opportunities. As we've seen, it takes vast and varied data sources to achieve these goals. What challenges do firms face as they navigate their large and complex data environments? How do they balance data quality and security while continuing to evolve and grow the business?

We compared Pacers, which we defined as respondents who indicated their organizations were effective and very effective at keeping pace with business data needs (59%), with the rest of respondents, which we categorized as Chasers (41%). We compared the priorities, challenges, and approaches of the Pacers versus the Chasers, and we found that:

› **Data quality, security, and privacy are paramount in governance.** Two themes were consistent across companies: data quality and security/privacy. Overall, more than 60% of organizations rated these as "very important" aspects of data governance strategies. By comparison, all other aspects are secondary, considered "very important" by 43% or less of respondents (see Figure 5). To ensure a comprehensive data integration strategy that meets governance expectations and requirements, data professionals must prioritize quality, security, and privacy rather than focus narrowly on data movement activities like extracting, mapping, and loading.

FIGURE 5
Data Quality And Security/Privacy Are Paramount In Governance



Base: 164 US and UK data governance decision-makers and influencers

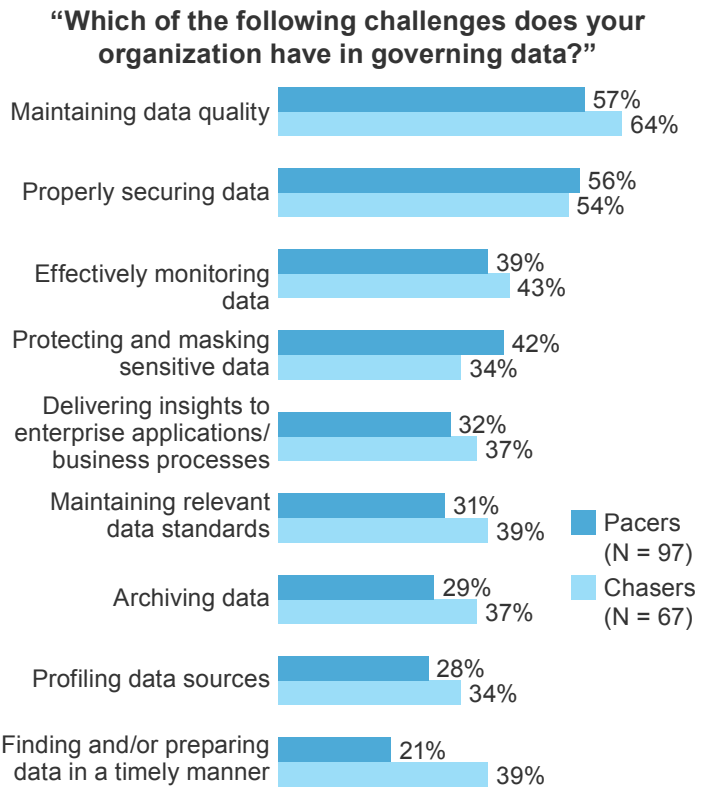
Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

› **At the same time, data quality and security top the list of challenges firms face.** Overall, 60% of survey respondents selected maintaining data quality as a top data governance challenge, and 55% percent selected properly securing data. Chasers indicated slightly more difficulty across nearly all challenges, particularly in finding and/or preparing data in a timely manner, which

they were nearly twice as likely to cite as Pacers (see Figure 6). This indicates that a lack of data governance, which includes better planning processes, creates productivity issues and ultimately affects business objectives.

FIGURE 6

Maintaining Quality And Securing Data Challenge Most Firms, But Chasers Face Further Challenges



Base: variable US and UK data governance decision-makers and influencers

Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

› **Amplification of data insights makes data quality an even bigger imperative.** Top data consumers are employees: end users (63%), line-of-business managers (62%), and executives (52%). The majority of these stakeholders access insights from data analytics environments weekly or more. But in addition to internal users, external stakeholders are also active data consumers. About a third of respondents indicated that their customers (36%), prospective customers (32%), partners (30%), and the general public (33%) access insights from their data environments at least weekly (see Figure 7). As internal and external parties access, interpret, and amplify insights, accurate and up-to-date data is crucial in order to make consistent and informed decisions. Security and privacy controls are also critical to maintain quality and ensure trust.

FIGURE 7

Data Insights Are Amplified Primarily Through Internal Users But Also Through External Stakeholders

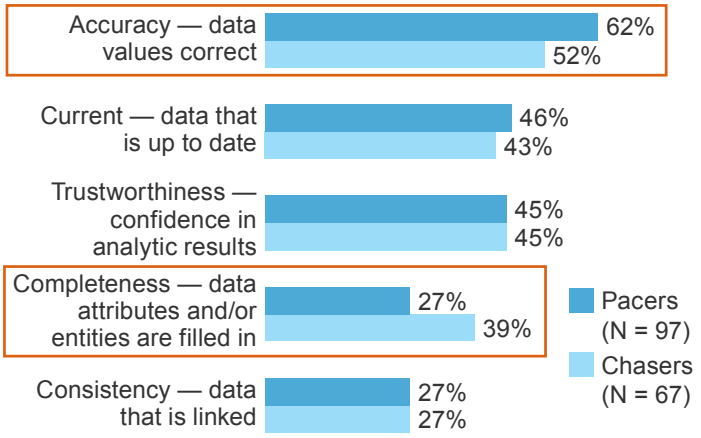
Base: 164 US and UK data governance decision-makers and influencers

Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

› **Organizations need to know the data is right.** Accuracy is the top-ranked aspect of data quality, with 58% of survey respondents who ranked it in their top three. Data must also be current and trustworthy. Chasers ranked completeness relatively high on the list: 39% ranked completeness in their top three versus 27% of Pacers (see Figure 8). Chasers can benefit from an architected integration approach by incorporating data governance controls as nascent aspects to offset issues that limit access and visibility into needed data.

FIGURE 8
Quality Hinges On Data Accuracy

"As it pertains to new data formats and data variety, what aspects of data quality are most important to your organization?"
(Percentage ranked in top three)



Base: variable US and UK data governance decision-makers and influencers (top five responses shown)

Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

Mature Organizations Use Data Governance For Competitive Advantage

Chasers — those less effective at keeping pace — will benefit from early involvement and collaboration with business stakeholders to maximize topline business outcomes while mitigating business risk and inefficiencies. Pacers demonstrate that piecing together data without business context limits possible uses of data, restricting data value to the business.

When we compared organizations that rated themselves high or very high in data governance maturity (64%) with those with lower maturity ratings, we found a stark contrast in data governance approaches and benefits. Our survey and Forrester's research reveal that:

› **Governance of data shows a spectrum for mature organizations.** Data types serve different purposes, and insights are amplified through various organizational stakeholders. Mature governance organizations take into account line-of-business involvement, policies, and standards in the context of various data types. For example, in mature governance organizations, IT ultimately owns big data, but cross-functional teams are responsible for setting governance policies and compliance. This works well given the technical requirements of managing big data. On the other hand, business owners, such as marketing leaders, own customer data (i.e., their 360-degree view of customers). They set policies for this data, but they leave governance execution and compliance to cross-functional teams. Less mature organizations do not apply this nuanced framework. IT teams are the de facto leads across data types (see Figure 9). By taking a more nuanced approach, organizations can apply appropriate levels of control for different data types (i.e., a spectrum), which allows certain analytics environments to be more nimble. For example, delivering governed data at scale in customer insights environments can help businesses quickly make decisions that ultimately improve the customer experience.

› **Ambiguous blending requirements are offset through IT/business collaboration and data certification.** Forrester's research highlights the challenges that organizations face in defining data quality programs, particularly in big data environments. Quality requirements

FIGURE 9

Mature Governance Organizations Involve Business Owners Strategically For Certain Data Types, While Others Rely On IT By Default

"For each of the following types of data, who in your organization is primarily responsible for each stage of data governance?"

Data Type	Governance Stage	Primary Owner — Mature Governance Orgs (N = 105)	Primary Owner — Less Mature Governance Orgs (N = 59)
Big data	Policy setting	Cross-functional data governance team	IT
	Executing governance programs	IT or cross-functional data governance team	IT
	Compliance	Cross-functional data governance team	IT
	Owning the data	IT	IT
Customer data (360-degree view of customers)	Policy setting	Line of business	IT or cross-functional data governance team
	Executing governance programs	Cross-functional data governance team	IT
	Compliance	Cross-functional data governance team	IT
	Owning the data	Line of business	IT or line of business

Base: variable US and UK data governance decision-makers and influencers

Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

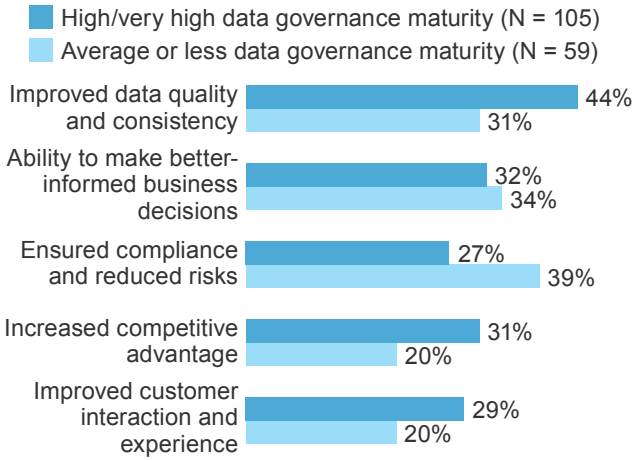
are difficult to define on the front end of data capture for environments that aggregate a variety of sources and/or include big data. To overcome, mature organizations allow analysts to play in the data first as a way to test and validate what is needed for integration. Additionally, by visualizing the data in the context of their use case, analysts can create more complete data quality standards that best fit the intended use and business outcome.[4]

› **Mature governance organizations reap quality and competitive benefits, while others primarily reduce risk.** In the past 12 months, organizations with high governance maturity have seen significant improvements in their ability to: 1) enhance data quality and consistency; 2) make better-informed business decisions; 3) increase competitive advantage; and 4) improve customer experience, effectively supporting topline business objectives. In contrast, less mature organizations still struggle with data quality and are primarily focusing on bottom-line benefits to ensure compliance and reduce risk (see Figure 10). As governance strategies mature, governance becomes a critical tool for organizations to harness data to drive key strategic and growth imperatives.

FIGURE 10

Mature Governance Organizations Reap Rewards Rather Than Primarily Reducing Risk

"In the past 12 months, where have you seen the most significant impact of your organization's data governance strategy?"
(Percentage ranked in top three)



Base: variable US and UK data governance decision-makers and influencers (top five benefits shown)

Source: A commissioned study conducted by Forrester Consulting on behalf of Pentaho, June 2015

Key Recommendations

Data integration and governance is more than the ability to extract, transform, and load data. Data professionals need to take on more than the system and syntactical view of data if they are going to not only meet, but also keep pace with, today's organizational demands for data. Developing a framework that describes the insights, decisions, consumption, and business processes that data supports ensures that data is not just available but also accurate and trusted. This turns data into a strategic tool to improve customer experience and address rising customer expectations. To succeed, data professionals need to expand the way they work with data, assess their systems, and include business owners in the process. Five recommendations will make this transition possible and successful:

- **Transition data integration to a strategic data competency by maturing data governance practices.** Data integration is often viewed as the bus that picks up and delivers data. Any transformations emphasize standardizing and mapping of data. Data quality and security/privacy priorities require data professionals to think ahead toward how data will be used and provide value to ensure that data integration efforts deliver accurate and relevant data for topline business objectives.
- **Tailor data governance to the context of data use.** Mature data governance organizations and those that keep pace with business demands for data recognize that there is no one-size-fits-all standard when it comes to data use. Different data types, sources, and uses demand a different set of rules, standards, and processes to govern the data. Data professionals need to create a framework that allows for a spectrum of data governance controls aimed at delivering business value.
- **Include business owners when it comes to data integration and governance.** Expanding the volumes and variety of data needed to gain better insights can leave data professionals struggling to put data requirements in place ahead of data demand. Business owners play a critical role in the process to profile data and evaluate data rules and standards. This improves IT's ability to deliver data services that are better aligned and overcome seeming hurdles of data ambiguity.
- **Embrace business-driven governance models to drive data flexibility and value.** Companies that primarily rely on IT teams to define and implement data governance should aspire to involve business owners over time. However, understand that this is a transition, during which flexibility is key. Involve business stakeholders on an informal basis to help define formal roles and responsibilities. From a technology perspective, pick flexible integration solutions that can support varied data sources and structures and evolve with business needs.
- **Adopt data integration solutions that support data governance upfront and deliver throughout.** Data governance requires solutions to execute rules and standards across various data types, sources, uses, and consumption points. Each step of the data pipeline requires tuning data governance for this spectrum. Data integration tools that simplify the development and oversight of processes for data professionals to deliver governed data will be critical.

Appendix A: Methodology

In this study, Forrester conducted an online survey of 164 business and IT leaders in the US and the UK with decision-making authority or influence over data governance at their organization. The survey targeted companies with 1,000-plus employees across all industries, but respondents from the healthcare, government, financial services, telecom, and energy/utilities verticals took priority. The survey evaluated data analytics objectives, governance challenges, and governance best practices. Survey participants included decision-makers in IT and line-of-business roles such as finance or operations, with IT comprising half the sample and line of business comprising the other half. Questions provided to the participants asked focused on data utilization, analytics capabilities, and data governance priorities, challenges, and frameworks. Respondents were offered a small incentive as a thank you for time spent on the survey. The study began and was completed in June 2015.

Appendix B: Supplemental Material

RELATED FORRESTER RESEARCH

“Data Quality Market Overview: Trust Your Data To Succeed With Customers,” Forrester Research, Inc., May 15, 2015

“Zombie Data May Be Taking Over Your Business,” Forrester Research, Inc., July 13, 2015

Appendix C: Endnotes

[1] These are distinct data sources, not number of extracts. For example, we asked respondents to consider extracts from their ERP system as a single data source, even if they pull feeds from multiple tables.

[2] Forrester defines the term "data hub" as a new data management architecture that employs a central low-cost big data management platform as a central landing, transformation, and analytic zone for data. Source: "Deliver On Big Data Potential With A Hub-And-Spoke Architecture," Forrester Research, Inc., June 12, 2013.

[3] Sixty-nine percent of respondents rated "address rising customer expectations" and 75% rated "improving customer experience" as a critical/high priority for their business in the next 12 months. Source: Global Business Technographics® Data And Analytics Survey, 2015, Forrester Research, Inc.

[4] Source: "Big Data Quality: Garbage In, Gold Out," Forrester Research, Inc., April 9, 2014.