



Dustin Smith, Product Manager

3 Shifts in the Modern Data Environment

And What it Means for IT Leaders

Table of Contents

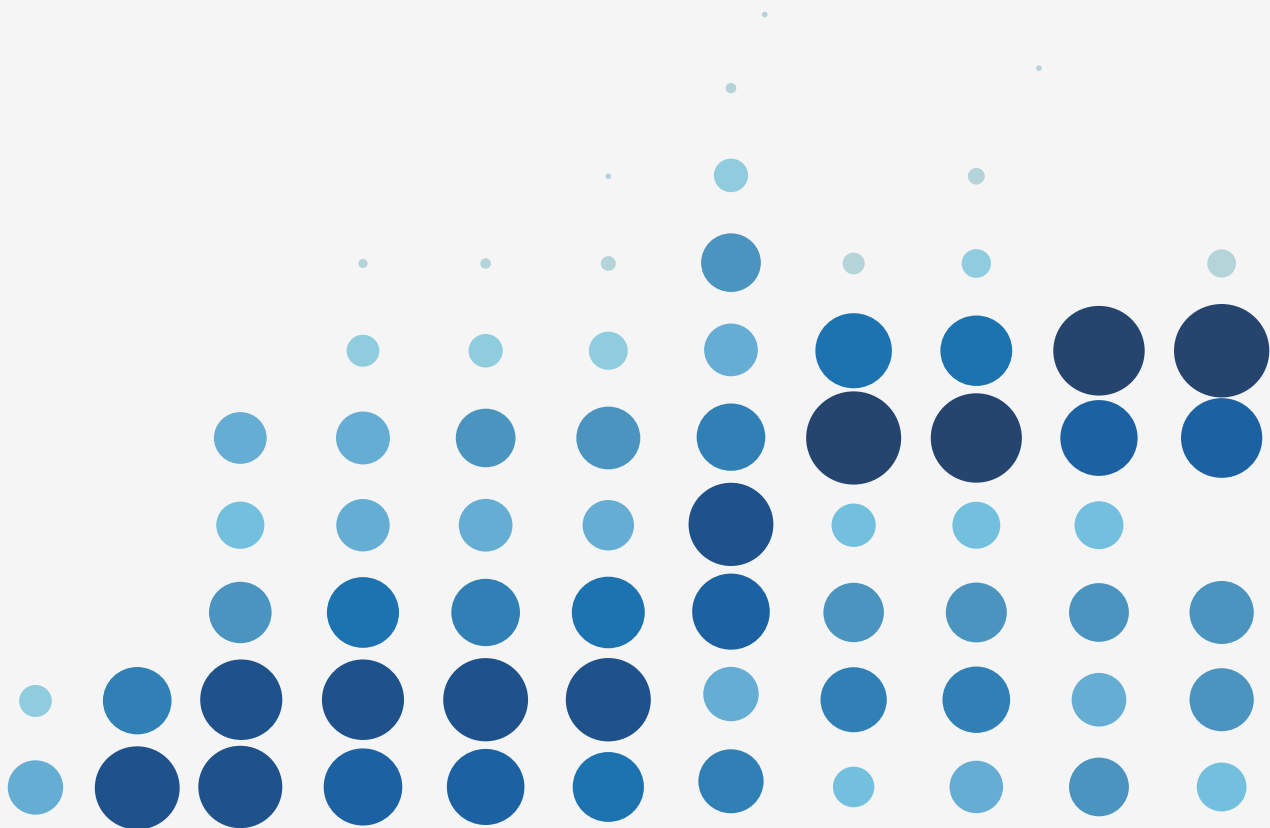
The Problem Isn't Different. Just Harder.....3

1. Think Pipelines, Not Buckets.....5

2. Use Need-Based Data Landing Zones.....7

3. Transforming from Data Protector to Data Mentor..... 11

The Problem Is Harder - But Not Impossible 13



The Problem Isn't Different. Just Harder.

Providing organizations with reliable data for better decision-making is an undertaking that has not fundamentally changed in decades. Despite massive technology advances and new tactics, the IT organization managing data infrastructure today still has the same overall mission: moving data from its moment of creation and making it accessible and understandable by decision-makers at the moment of need.

However, while the objective has stayed the same, the obstacles to successfully create and maintain a source of analytical truth within a business have become exponentially more difficult.

Perhaps the biggest hurdle in recent years within the modern data environment has been new sources of data that generate unprecedented amounts of output, often with very little (if any) structure. From clickstreams, server logs, and social media sources to machine and sensor readings, the onslaught of data from these channels has been overwhelming—literally. From an economic and performance point of view, traditional enterprise data warehouses (EDWs) simply cannot keep up with this data tidal wave.

This has sparked a complete re-think of data capture and analysis strategies and given rise to a new generation of data storage solutions aimed at schema-less capture, hardware scalability, and the moving of compute capability closer to (if not on top of) data stores themselves.

Though still young by relational database standards, these newer, non-relational solutions have gained serious traction in recent years and matured rapidly to support some of the largest and most complex corporate enterprises in the world. While this has been done largely as a means to complement existing enterprise data warehouse infrastructures, it never the less creates a more complex data ecosystem for IT to manage.

Adding to the hurdles IT must overcome in the ongoing mission to maintain a healthy data environment is the availability of data from cloud applications. Many organizations use applications like Google Analytics, Salesforce, Netsuite, Zendesk, and others as core parts of their infrastructure. The data they generate is critical to organizational reporting. Integrating data from these cloud solutions and making it accessible to the company has become a standard requirement for IT.

With the traditional EDW no longer functioning as the sole data destination, the question of “when, where, how, and if” to bring cloud application data into the corporate data environment is an ongoing and heated discussion.

Lastly, the emerging trend of self-service analytics for organizations of all sizes means that more and more non-technical users (no formal IT/data training) are doing both data discovery and reporting. Businesses embracing this movement often see a dramatic reduction in (if not full elimination of) IT’s responsibilities for producing analytics. While this shift is critical to the overall success of an organization adopting a data-driven mindset, it puts new pressure on IT groups to ensure data governance as well as provide data analytics training and technology.

In an effort to meet these new challenges, many IT organizations rush to adopt new technologies and tactics, but fail to see how these hurdles have actually shifted the way information technology groups need to approach the goal of managing data from “creation to consumption.” Big Data solutions, cloud data integration, and self-service analytics are all answers to bigger technology problems, but in order to deploy them effectively in an organization, the IT playbook needs to change.

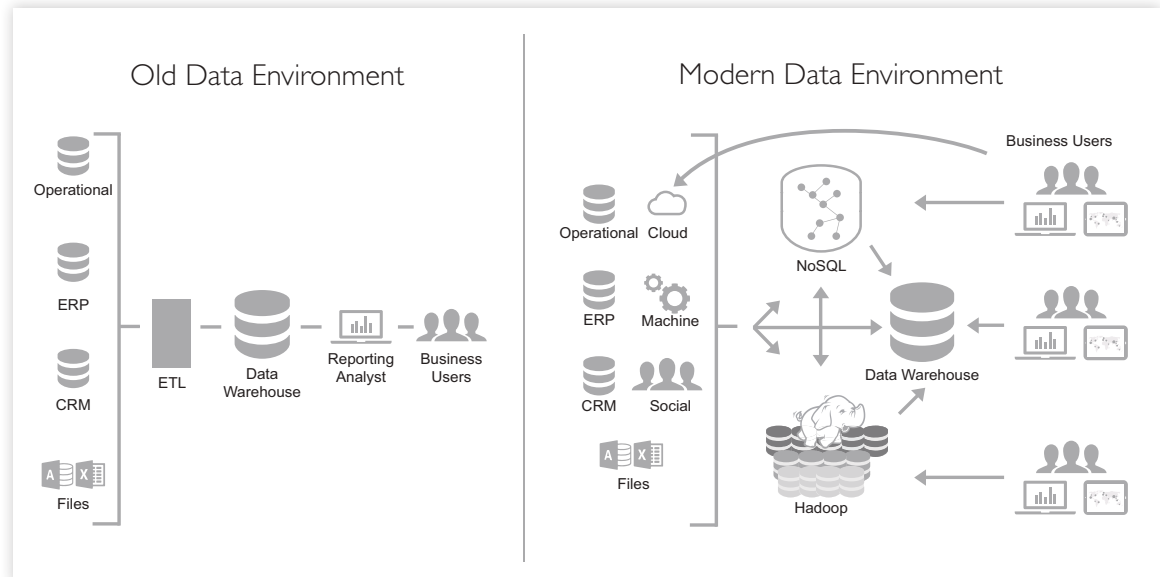
This paper seeks to outline three major thought shifts concerning the modern data environment that Information Technology leaders need to understand in order to support data-driven decision making within their organization.



Think Pipelines, Not Buckets.

The enterprise data warehouse is not dead. It just has more friends. From Hadoop clusters to NoSQL databases, the relational EDW is no longer the only officially-sanctioned place for data to reside.

This means that the data movement process no longer needs to be centralized around a single location. In fact today a modern data platform is something more like the connective neural tissue joining the hemispheres of an organization's 'data brain.' Combine this with the now-ubiquitous cloud solutions that allow for infrastructure and services to spin up pipeline/ETL projects in hours, and what you have is a recipe for moving data around an organization like never before.



Unfortunately, this opportunity is largely missed by many IT groups as many organization's data management handbook was written using yester-year's "single-bucket-of-data" mindset.

The trick to shifting from a "buckets" to "pipelines" mentality is accepting that not all data questions within an organization can be answered from any one data source. An EDW may not have the capacity to allow the data granularity required to drill into five years' worth of emails scraped directly from message server logs. And a Hadoop deployment may not be able to provide sub-second query responses against sales data needed for real-time analysis. In the real world, many business problems require both data granularity and query speed from one or multiple sources—but at different periods and varying successions to complete a data project.

Given these requirements, the question for IT is no longer “Which bucket does this data go in?” Rather, the question is: “What stage of data is necessary?” and “How do I let users easily move between both?”

This is ultimately a pipeline discussion since the actual location of any given source of data may change as the project moves through different stages of completion. The pipeline mentality also extends much further than just the traditional data integration world, moving into the realm of organizational workflow.

How are business users arriving at their data answers? Most IT groups know that there are individual, unsanctioned data repositories being kept in spreadsheets across the organization. But IT has felt powerless to contain the spread of siloed, ungoverned, local file data management. People on the business side often feel that working around the corporate data environment is the only practical solution—so that’s what they do.

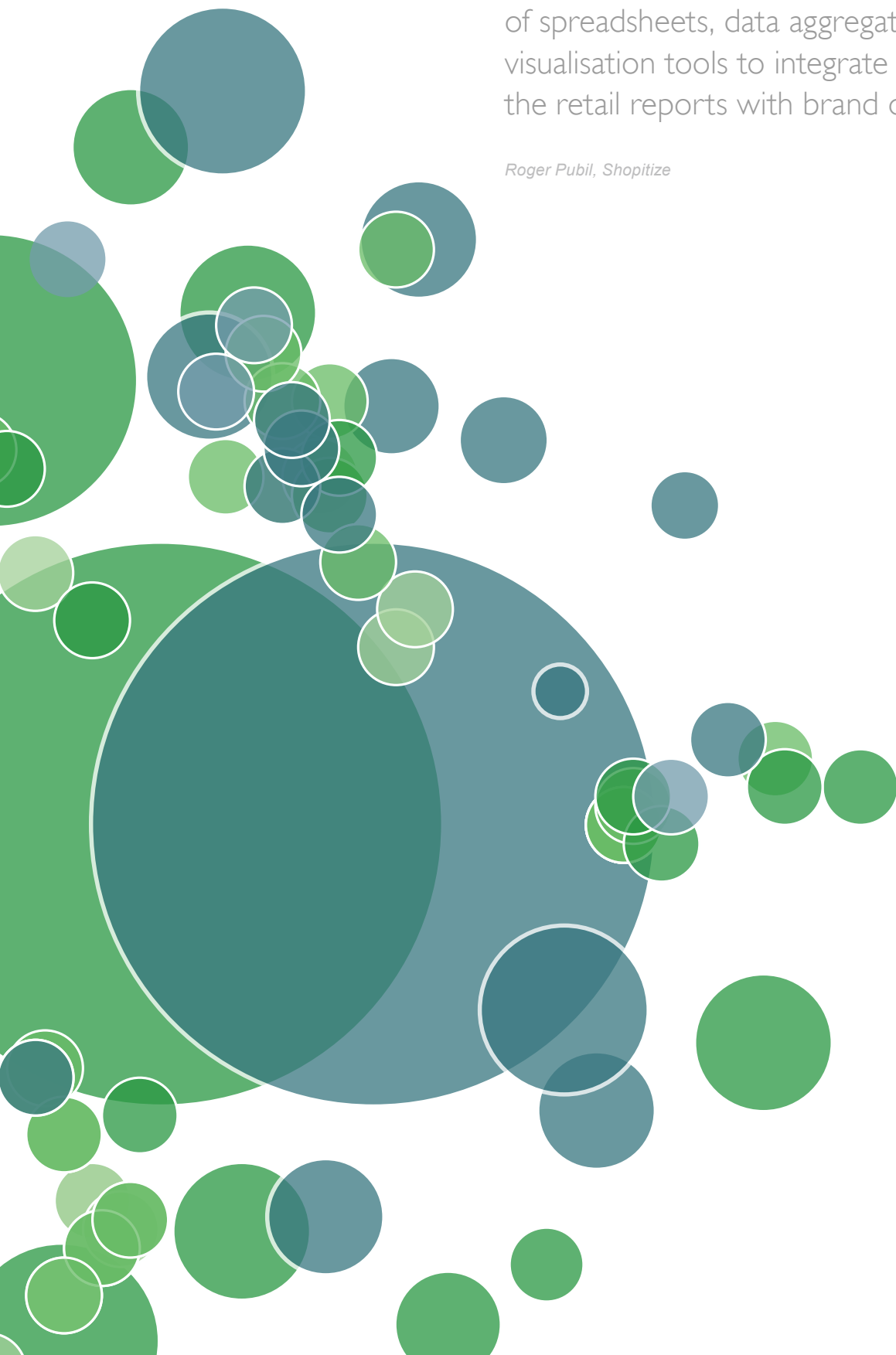
When IT groups understand the full arc of how business users find, clean, analyze, and present data—and become enablers in the process (while simultaneously coaching governance) —only then does the shift to a pipeline mentality occur. Embracing the idea that the movement of data, both in the traditional batch IT sense as well as in the ad hoc query nature of business users, is more important than the fortified data bunker represents a pivotal and critical shift for technology groups.





We previously relied on a mixed bag of spreadsheets, data aggregation, and visualisation tools to integrate and present the retail reports with brand owners...

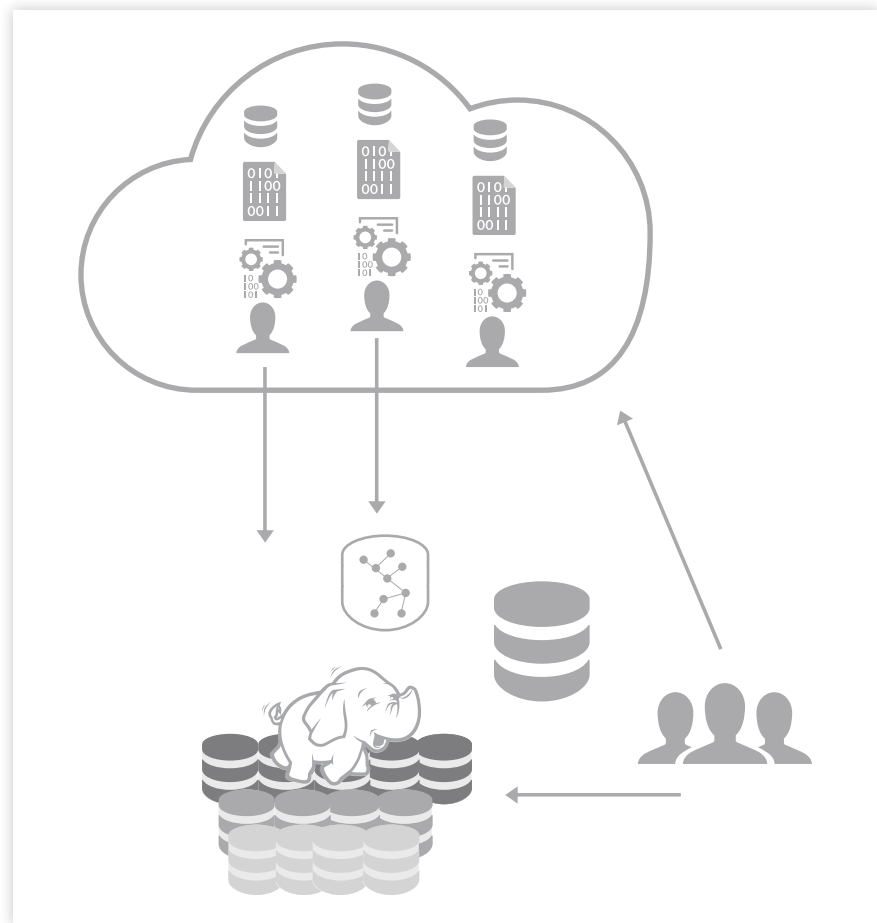
Roger Pubil, Shopitize



2.

Use Need-Based Data Landing Zones

Cloud application data is not always meant to land directly in the enterprise data warehouse. The decision should always map back to the needs of the organization. Or, more clearly, the decision should map back to what is necessary to make the data useful to the organization.



A common misconception among IT groups, even those that have accepted the “no-single-data-bucket” approach, is that web application data should automatically be integrated directly into the EDW. This makes sense on the surface, given that many successful business cloud applications are integral parts of sales, marketing, and support infrastructure. This is highly-actionable, highly-accessed, often well-structured data. The ROI of placing these assets in an EDW, where price per byte is highest, appears sound.

However, the critical questions IT groups need to be asking are:

- “Is the data ready for analysis straight from the cloud?” In other words, “Is it clean, and should we bother moving it?”
- “Is the real value of the cloud data in question only realized after it is combined with additional data?”

Depending on the answers to these questions, where an IT group lands its cloud data assets (if they choose to bring them on-premise at all) might change from organization to organization and from application to application.

Any IT group in charge of a decent-sized Salesforce.com deployment will tell you immediately that allowing un-deduplicated opportunity records into the EDW is a recipe for disaster. From inaccurate commission payouts to grossly over-projected revenue numbers, these problems represent just the tip of the iceberg when it comes to dealing with improperly cleansed sales data.

Likewise, website and marketing data from cloud sources like Google Analytics, Eloqua, and Marketo can play havoc on an organization’s ability to track lead flow and calculate acquisition costs for customers if the records’ accuracy can’t be trusted.

This accuracy problem gets exponentially worse when the sources must be merged. This is often the case when an organization needs to paint a complete picture of its customer funnel—from first website hit down to sales decision. While nearly all IT groups understand the importance of clean data—cloud or other—they often don’t think about it in the context of where it lands as a first stop when being pulled down from a web application.

An IT organization with a multi-bucket/pipeline-focused data environment will evaluate cloud data integration on the basis of the data asset’s value to the business at the time it is brought on-premise. Where a cloud data source is relatively clean and provides an extremely high amount of value to the business upon arrival, IT groups are well advised to land it where it can be accessed most rapidly (relational EDW).

However, when dealing with web application data that requires heavy and/or complex processing prior to being widely trusted by the business, IT groups can take advantage of high-compute, low cost-per-byte environments such as a Hadoop environment. With this second approach, resources can be maximized on data cleaning and transformation without affecting the speed of the EDW. From there, IT working with the business can decide whether the cleansed data should be moved into an EDW and/or accessed directly within the larger environment.

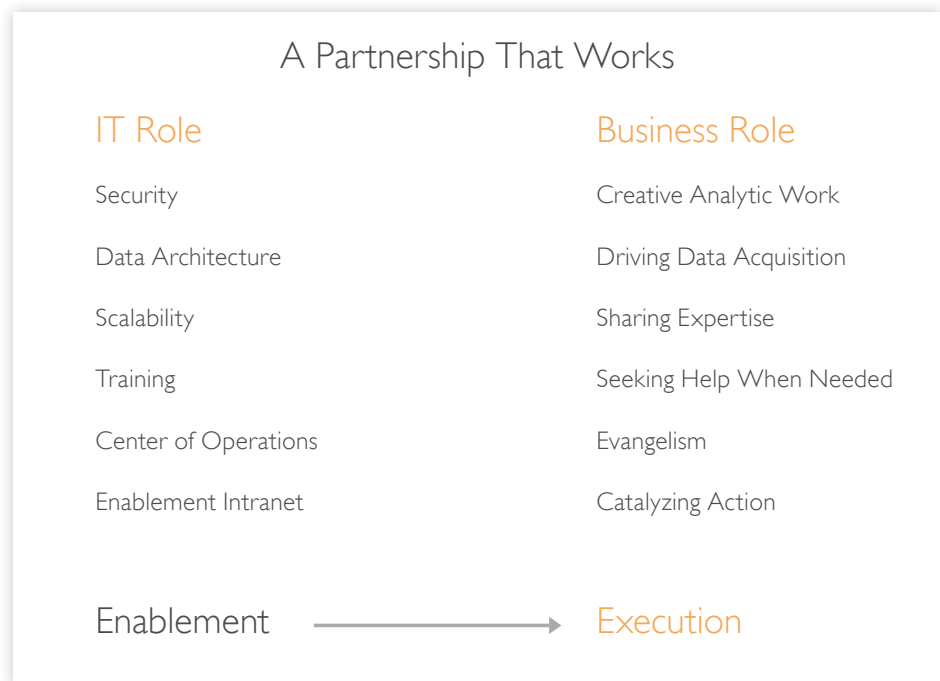
Lastly, it may not be necessary to move cloud application data at all. It is increasingly common that web applications are providing easy access points for their backend repositories meaning business users have the capability to use self-service reporting and analytics tools to do their own investigations against live data. For IT organizations that are still interested in having a middle layer of authorization and governance in these scenarios, some of the business friendly self-service analytics solutions also allow for a proxy connection scenario. This can solve IT's need for everything from basic user access to highly involved business logic. Given this reality, the first question regarding cloud application data integration is not "where," but "should." Is there a compelling need to move the data at all (cleansing, additional value, etc.) or can it be left in place and potentially integrated at the business user level?

3.

Transforming from Data Protector to Data Mentor

The self-service analytics movement within companies should be considered an offer of a handshake to IT, not a threat of hostile takeover. And those IT groups that extend their hands first stand to win everything.

In the same way that the modern data environment has expanded to incorporate more than just relational data stores, so too has the analytics environment grown to include tools that give tremendous value to everyday business users. This has sparked a massive departure from the traditional process in which all analytics are funneled through a highly-trained few.



As a result, forward-thinking IT groups are morphing into data and analytics thought leaders within their larger organizations and redefining the way they work with business teams to solve data challenges. This shift in the IT mindset from “data protector” to “data mentor” is perhaps the most significant shift within the modern data environment. This is a key factor in an organization’s success or failure in becoming truly data-driven.

Redefining this relationship between IT and the business means adjusting the default data governance attitude from “no” to “yes.” Traditionally, a business user was automatically denied access to data unless specifically given approval. This new model assumes that any business user anywhere in the company can access any data asset so long as it does not violate compliance regulations.

The power of this attitude change is massive. The impact can be seen in the way that business users begin to seek new data to tackle new problems. Their willingness to do so is directly related to their perception of how much effort accessing new data will take. When armed with an easy-to-use, self-service analytics solution and encouraged by the knowledge that their IT group is supportive of widespread data access, people in the business can become change agents. And these agents are markers of successful entry into the early stages of becoming a data driven company.

Simply opening up access to data and handing out new software does not solve the bigger problem of actually helping the business make better decisions with data though. IT organizations that have successfully partnered to transition their companies into data-driven decision engines have done so by truly embracing the role of data mentor. Often those IT team members that previously controlled the analytics process become the guides for the rest of the company. They have the skills to train others in critical functions such as proper data analysis and reporting. Educating business users in this way effectively scales the analytical reporting power of the entire organization by an order of magnitude. This also helps organizations to achieve higher and faster returns on their investments in self-service analytics solutions. Most importantly though, using data to better understand the business as a whole becomes part of the wider organization’s operating DNA.

The Problem Is Harder - But Not Impossible

From dealing with wholly new data storage and processing solutions to understanding when, where, and if cloud data should enter the ecosystem, IT groups' responsibilities for managing organizational data flow are rapidly becoming more difficult. Add to this the self-service analytics movement and these shifts become something bigger altogether—a sea change in the way companies deal with data.

Organizations need IT to guide them through this change. Much of the uncertainty that IT groups feel is largely due to the understanding that the issue is no longer about fitting new technology into existing business processes. Big Data, cloud technologies, and enabling a self-service analytics model represent fundamental changes in how IT groups need to approach the modern data environment.

For those organizations that have already started re-evaluating their mentality towards data and analytics in order to accommodate these shifts, the hurdles are real—but so are the discoveries. Change is hard both for business users and for IT. When redefining the relationship between the two, there will be elements from each camp that will resist. But there are also those that will rise up to embrace it.

As the rigid walls and processes around organizational data access change, IT groups will find unexpected analytics champions in every corner of their company who are eager to be the vanguard of a new data driven company. The best IT leaders will pair these new champions with their seasoned data professionals as mentors. The champions will learn their way through the evolving data ecosystem and IT will learn the business problems people are trying to solve. These first steps lay the groundwork for a culture of analytics and data discovery within a company where decisions are powered with the confidence and accuracy of the modern data environment.

Your Next Step

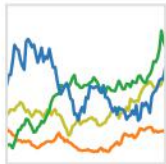
This paper outlines the shifts occurring in the modern data environment. To learn about implementing a modern approach to analytics that capitalizes on these shifts, read on:

→ [Website: *The Modern Approach to Enterprise Analytics*](#)

→ [Whitepaper: *The Road Forward - An IT Roadmap for Scaling Self-Service Reporting*](#)

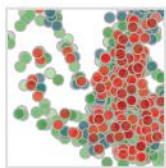
About Tableau

Tableau helps people see and understand data. Tableau helps anyone quickly analyze, visualize and share information. More than 21,000 customer accounts get rapid results with Tableau in the office and on-the-go. And tens of thousands of people use Tableau Public to share data in their blogs and websites. See how Tableau can help you at www.tableau.com.



Additional Resources

[Start a Free Trial](#)



Explore Other Resources

- [Product Demo](#)
- [Training & Tutorials](#)
- [Community & Support](#)
- [Customer Stories](#)
- [Solutions](#)

