



STRATEGY

What's Your Data Strategy?

by Leandro DalleMule and Thomas H. Davenport

FROM THE MAY-JUNE 2017 ISSUE

More than ever, the ability to manage torrents of data is critical to a company's success. But even with the emergence of data-management functions and chief data officers (CDOs), most companies remain badly behind the curve. Cross-industry studies show that on average, less than half of an organization's structured data is actively used in making decisions—and less than 1% of its unstructured data is analyzed or used at all. More than 70% of employees have access to data they should not, and 80% of analysts' time is spent simply discovering and preparing data. Data breaches are common, rogue data sets propagate in silos, and companies' data technology often isn't up to the demands put on it.

Having a CDO and a data-management function is a start, but neither can be fully effective in the absence of a coherent strategy for organizing, governing, analyzing, and deploying an organization's information assets. Indeed, without such strategic management many companies struggle to protect and leverage their data—and CDOs' tenures are often difficult and short (just 2.4 years on average, according to Gartner). In this article we describe a new framework for building a robust data strategy that can be applied across industries and levels of data maturity. The framework draws on our implementation experience at the global insurer AIG (where DalleMule is the CDO) and our study of half a dozen other large companies where its elements

have been applied. The strategy enables superior data management and analytics—essential capabilities that support managerial decision making and ultimately enhance financial performance.

The “plumbing” aspects of data management may not be as sexy as the predictive models and colorful dashboards they produce, but they’re vital to high performance. As such, they’re not just the concern of the CIO and the CDO; ensuring smart data management is the responsibility of all C-suite executives, starting with the CEO.

Defense Versus Offense

Our framework addresses two key issues: It helps companies clarify the primary purpose of their data, and it guides them in strategic data management. Unlike other approaches we’ve seen, ours requires companies to make considered trade-offs between “defensive” and “offensive” uses of data and between control and flexibility in its use, as we describe below. Although information on enterprise data management is abundant, much of it is technical and focused on governance, best practices, tools, and the like. Few if any data-management frameworks are as business-focused as ours: It not only promotes the efficient use of data and allocation of resources but also helps companies design their data-management activities to support their overall strategy.

Data defense and offense are differentiated by distinct business objectives and the activities designed to address them. Data defense is about minimizing downside risk. Activities include ensuring compliance with regulations (such as rules governing data privacy and the integrity of financial reports), using analytics to detect and limit fraud, and building systems to prevent theft. Defensive efforts also ensure the integrity of data flowing through a company’s internal systems by identifying, standardizing, and governing authoritative data sources, such as fundamental customer and supplier information or sales data, in a “single source of truth.” Data offense focuses on supporting business objectives such as increasing revenue, profitability, and customer satisfaction. It typically includes activities that generate customer insights (data analysis and modeling, for example) or integrate disparate customer and market data to support managerial decision making through, for instance, interactive dashboards.

Offensive activities tend to be most relevant for customer-focused business functions such as sales and marketing and are often more real-time than is defensive work, with its concentration on legal, financial, compliance, and IT concerns. (An exception would be data fraud protection, in which seconds count and real-time analytics smarts are critical.) Every company needs both

offense and defense to succeed, but getting the balance right is tricky. In every organization we've talked with, the two compete fiercely for finite resources, funding, and people. As we shall see, putting equal emphasis on the two is optimal for some companies. But for many others it's wiser to favor one or the other.

Some company or environmental factors may influence the direction of data strategy: Strong regulation in an industry (financial services or health care, for example) would move the organization toward defense; strong competition for customers would shift it toward offense. The challenge for CDOs and the rest of the C-suite is to establish the appropriate trade-offs between defense and offense and to ensure the best balance in support of the company's overall strategy.

Decisions about these trade-offs are rooted in the fundamental dichotomy between standardizing data and keeping it more flexible. The more uniform data is, the easier it becomes to execute defensive processes, such as complying with regulatory requirements and implementing data-access controls. The more flexible data is—that is, the more readily it can be transformed or interpreted to meet specific business needs—the more useful it is in offense. Balancing offense and defense, then, requires balancing data control and flexibility, as we will describe.

Single Source, Multiple Versions

Before we explore the framework, it's important to distinguish between information and data and to differentiate information architecture from data architecture. According to Peter Drucker, information is “data endowed with relevance and purpose.” Raw data, such as customer retention rates, sales figures, and supply costs, is of limited value until it has been integrated with other data and transformed into information that can guide decision making. Sales figures put into a historical or a market context suddenly have meaning—they may be climbing or falling relative to benchmarks or in response to a specific strategy.

A company's data architecture describes how data is collected, stored, transformed, distributed, and consumed. It includes the rules governing structured formats, such as databases and file systems, and the systems for connecting data with the business processes that consume it. Information architecture governs the processes and rules that convert data into useful information. For example, data architecture might feed raw daily advertising and sales data into information architecture systems, such as marketing dashboards, where it is integrated and analyzed to reveal relationships between ad spend and sales by channel and region.

Many organizations have attempted to create highly centralized, control-oriented approaches to data and information architectures. Previously known as information engineering and now as master data management, these top-down approaches are often not well suited to supporting a broad data strategy. Although they are effective for standardizing enterprise data, they can inhibit flexibility, making it harder to customize data or transform it into information that can be applied strategically. In our experience, a more flexible and realistic approach to data and information architectures involves both a single source of truth (SSOT) and multiple versions of the truth (MVOTs). The SSOT works at the data level; MVOTs support the management of information.

In the organizations we've studied, the concept of a single version of truth—for example, one inviolable primary source of revenue data—is fully grasped and accepted by IT and across the business. However, the idea that a single source can feed multiple versions of the truth (such as revenue figures that differ according to users' needs) is not well understood, commonly articulated, or, in general, properly executed.

The key innovation of our framework is this: It requires flexible data and information architectures that permit both single and multiple versions of the truth to support a defensive-offensive approach to data strategy.

The Elements of Data Strategy

	DEFENSE	OFFENSE
KEY OBJECTIVES	Ensure data security, privacy, integrity, quality, regulatory compliance, and governance	Improve competitive position and profitability
CORE ACTIVITIES	Optimize data extraction, standardization, storage, and access	Optimize data analytics, modeling, visualization, transformation, and enrichment
DATA-MANAGEMENT ORIENTATION	Control	Flexibility
ENABLING ARCHITECTURE	SSOT (Single source of truth)	MVOTs (Multiple versions of the truth)

The SSOT is a logical, often virtual and cloud-based repository that contains one authoritative copy of all crucial data, such as customer, supplier, and product details. It must have robust data provenance and governance controls to ensure that the data can be relied on in defensive and offensive activities, and it must use a common language—not one that is specific to a particular business unit or function. Thus, for example, revenue is reported, customers are defined, and products are classified in a single, unchanging, agreed-upon way within the SSOT.

Not having an SSOT can lead to chaos. One large industrial company we studied had more than a dozen data sources containing similar supplier information, such as name and address. But the content was slightly different in each source. For example, one source identified a supplier as Acme; another called it Acme, Inc.; and a third labeled it ACME Corp. Meanwhile, various functions within the company were relying on differing data sources; often the functions weren't even aware that alternative sources existed. Human beings might be able to untangle such problems (though it would be labor-intensive), but traditional IT systems can't, so the company couldn't truly understand its relationship with the supplier. Fortunately, artificial intelligence tools that can sift through such data chaos to assemble an SSOT are becoming available. The industrial company ultimately tapped one and saved substantial IT costs by shutting down redundant systems. The SSOT allowed managers to identify suppliers that were selling to multiple business units within the company and to negotiate discounts. In the first year, having an SSOT yielded \$75 million in benefits.

A New Data Architecture Can Pay for Itself

When companies lack a robust SSOT-MVOTs data architecture, teams across the organization may create and store the data they need in siloed repositories that vary in depth, breadth, and formatting. Their data management is often done in isolation with inconsistent requirements. The process is inefficient and expensive and can result in the proliferation of multiple uncontrolled versions of the truth that aren't effectively reused. Because SSOTs and MVOTs concentrate, standardize, and streamline data-sourcing activities, they can dramatically cut operational costs.

An SSOT is the source from which multiple versions of the truth are developed. MVOTs result from the business-specific transformation of data into information—data imbued with “relevance and purpose.” Thus, as various groups within units or functions transform, label, and report data, they create distinct, controlled versions of the truth that, when queried, yield consistent, customized responses according to the groups’ predetermined requirements.

One large financial services company doing business in more than 200 countries consolidated nearly 130 authoritative data sources, with trillions of records, into an SSOT. This allowed the company to rationalize its key data systems; eliminate much supporting IT infrastructure, such as databases and servers; and cut operating expenses by automating previously manual data consolidation. The automation alone yielded a 190% return on investment with a two-year payback time. Many companies will find that they can fund their entire data management programs, including staff salaries and technology costs, from the savings realized by consolidating data sources and decommissioning legacy systems.

The CDO and the data-management function should be fully responsible for building and operating the SSOT structure and using the savings it generates to fund the company's data program. Most important is to ensure at the outset that the SSOT addresses broad, high-priority business needs, such as applications that benefit customers or generate revenue, so that the project quickly yields results and savings—which encourages organization-wide buy-in.

Consider how a supplier might classify its clients Bayer and Apple according to industry. At the SSOT level these companies belong, respectively, to chemicals/pharmaceuticals and consumer electronics, and all data about the supplier's relationship with them, such as commercial transactions and market information, would be mapped accordingly. In the absence of MVOTs, the same would be true for all organizational purposes. But such broad industry classifications may be of little use to sales, for example, where a more practical version of the truth would classify Apple as a mobile phone or a laptop company, depending on which division sales was interacting with. Similarly, Bayer might be more usefully classified as a drug or a pesticide company for the purposes of competitive analysis. In short, multiple versions of the truth, derived from a common SSOT, support superior decision making.

A company's position on the offense-defense spectrum is rarely static.

At a global asset management company we studied, the marketing and finance departments both produced monthly reports on television ad spending—MVOTs derived from a common SSOT. Marketing, interested in analyzing advertising effectiveness, reported on spending after ads had aired. Finance, focusing on cash flow, captured spending when invoices were paid. The reports therefore contained different numbers, but each represented an accurate version of the truth.

Procter & Gamble has adopted a similar approach to data management. The company long had a centralized SSOT for all product and customer data, and other versions of data weren't allowed. But CDO Guy Peri and his team realized that the various business units had valid needs for customized interpretations of the data. The units are now permitted to create controlled data transformations for reporting that can be reliably mapped back to the SSOT. Thus the MVOTs diverge from the SSOT in consistent ways, and their provenance is clear.

Good Governance, Good Data

A sound data strategy requires that the data contained in a company's single source of truth (SSOT) is of high quality, granular, and standardized, and that multiple versions of the truth (MVOTs) are carefully controlled and derived from the same SSOT. This necessitates good governance for both data and technology. In the absence of proper governance, some common problems arise:

Data definitions may be ambiguous and mutable.

With no concrete definition at the outset of what constitutes the "truth" (whether an SSOT or MVOTs), stakeholders will squander time and resources as they try to manage nonstandardized data.

Data rules are vague or inconsistently applied.

If rules for aggregating, integrating, and transforming data are unclear, misunderstood, or simply not followed—particularly when data transformation involves multiple poorly defined steps—it's difficult to reliably replicate transformations and leverage information across the organization.

Feedback loops for improving data transformation are absent.

In its application of the SSOT-MVOTs model, the Canadian Imperial Bank of Commerce (CIBC) automated processes to ensure that enterprise source data and data transformations remained aligned. CIBC's CDO, Jose Ribau, explains that the company's SSOT contains all basic client profile and preference data; MVOTs for loan origination and customer relationship management transform the source data into information that supports regulatory reporting and improved customer experience. Automated synchronization programs connect SSOT and MVOTs data, with nightly "exception handling" to identify and address data-integrity issues such as inconsistent customer profiles.

Although the SSOT-MVOTs model is conceptually straightforward, it requires robust data controls, standards, governance, and technology. Ideally, senior executives will actively participate on data governance boards and committees. But data governance isn't particularly fun. Typically, enterprise CDOs and CTOs lead data and technology governance processes, and business and technology managers in functions and units are the primary participants. What's critical is that single sources

Complex data analyses such as predictive modeling may be undertaken by one group but prove useful across an organization. Without mechanisms for making these outputs available to others (by, for example, integrating them into appropriate MVOTs), stakeholders may needlessly duplicate work or miss opportunities.

Strong data governance usually involves standing committees or review boards composed of business and technology executives, but it relies heavily on robust technology oversight. If technology rules prevent a marketing executive from buying a server on his or her corporate purchasing card, it's much less likely that marketing will, for instance, create unregulated "shadow" MVOTs or a marketing analytic that duplicates an existing one.

of the truth remain unique and valid, and that multiple versions of the truth diverge from the original source only in carefully controlled ways.

A Lake of Data

Until a few years ago, technological limitations made it hard to build the SSOT-MVOTs data architecture needed to support a robust data strategy. Companies depended on traditional data warehouses that stored structured enterprise data in hierarchical files and folders, but these were not always suited to managing vast and growing volumes of data and new formats. To meet the need for a cheaper, more agile and scalable architecture, Silicon Valley engineers devised the "data lake," which can store virtually unlimited amounts of structured and unstructured data, from databases to spreadsheets to free text and image files. Data lakes are an ideal platform for SSOT-MVOTs architecture. A lake can house the SSOT, extracting, storing, and providing access to the organization's most granular data down to the level of individual transactions. And it can support the aggregation of SSOT data in nearly infinite ways in MVOTs that also reside in the lake. Data warehouses still have their uses: They store data for production applications (such as general ledger and order-management systems) that require tight security and access.

Striking a Balance

Let's return now to data strategy—striking the best balance between defense and offense and between control and flexibility. Whereas the CEO—often with the CIO—is ultimately responsible for a company's data strategy, the CDO commonly conceives it and leads its development and

execution. The CDO must determine the right trade-offs while dynamically adjusting the balance by leveraging the SSOT and MVOTs architectures.

It's rare to find an organization—especially a large, complex one—in which data is both tightly controlled and flexibly used. With few exceptions, CDOs find that their best data strategy emphasizes either defense and control (which depends on a robust SSOT) or offense and flexibility (enabled by MVOTs). Devoting equal attention to offense and defense is sometimes optimal, but in general it's unwise to default to a 50/50 split rather than making considered, strategic trade-offs. To determine a company's current and desired positions on the offense-defense spectrum, the CDO must bear in mind, among other things, the company's overall strategy, its regulatory environment, the data capabilities of its competitors, the maturity of its data-management practices, and the size of its data budget. For example, insurance and financial services companies typically operate in heavily regulated environments, which argues for an emphasis on data defense. (That is the case at AIG.) Retailers, operating in a less-regulated environment where intense competition requires robust customer analytics, might emphasize offense.

As Peri points out, defense and offense often require differing approaches from IT and the data-management organization. Defense, he argues, is day-to-day and operational, and at P&G is largely overseen by permanent IT teams focused on master data management, information security, and so forth. Offense involves partnering with business leaders on tactical and strategic initiatives. Leaders may be reluctant to engage with master data management, but they are happy to collaborate on optimizing marketing and trade promotion spending.

Of course, plenty of cases don't fall neatly into either the offense or the defense category: The CDO of a large hedge fund told us that he was less concerned with data protection than with rapidly gathering and using new data. The most valuable data for his fund is primarily external, publicly or commercially available, captured in real time, and already of good quality, structured,

and cleansed. Additionally, although his business is in financial services, it's not heavily regulated. Thus he focuses primarily on data offense. Wells Fargo's CDO, A. Charles Thomas, has enterprise responsibility for customer-related analytics, an offensive activity, and strives to keep the balance between offensive and defensive activities around 50/50, even structuring meeting agendas to focus equally on the two.

The tool “Assess Your Strategy Position” offers diagnostic questions that can help CDOs place their companies on the offense-defense spectrum and gauge whether their data strategy aligns with their corporate strategy. Determining an organization’s current and desired positions on the spectrum will force executives to make trade-offs between offensive and defensive investments. Of course, this tool is not a precise measure. CDOs should use the results to inform data strategy and discussions with other C-level executives.

Assess Your Strategy Position

Please enable JavaScript to view this assessment.

We find that companies with the most-advanced data strategies started at one point and gradually migrated to a new, stable position. For example, they may have shifted their focus from defense and data control toward offense as their data defense matured or competition heated up. The opposite path—from offense toward defense, and from flexible toward controlled—is possible but usually more difficult.

Consider how data strategy has shifted at CIBC. The bank established the chief data officer role a few years ago and for the first 18 months maintained a 90% defensive orientation, focusing on governance, data standardization, and building new data-storage capabilities. When Jose Ribau took over as CDO, in 2015, he determined that CIBC’s defense was sufficiently solid that he could

shift toward offense, including more-advanced data modeling and data science work. Today CIBC's data strategy strikes a 50/50 balance. Ribau expects that the new attention to offense will drive increased ROI from data products and services and nurture analytical talent for the future.

Regardless of what industry a company is in, its position on the offense-defense spectrum is rarely static. As competitive pressure mounts, an insurer may decide to increase its focus on offensive activities. A hedge fund may find itself in a tougher regulatory environment that requires rebalancing its data strategy toward defense. How a company's data strategy changes in direction and velocity will be a function of its overall strategy, culture, competition, and market.

Organizing Data Management

As with most organizational design, data-management functions can be built centrally or decentralized by function or business unit. The optimal design will depend on a company's position on the offense-defense spectrum. A centralized data function typically has a single CDO with accountability across the entire organization, ensuring that data policies, governance, and standards are applied consistently. This design is most suitable for businesses that focus on data defense.

Conversely, several companies we studied found that data offense can be better executed through decentralized data management, typically with a CDO for each business unit and most corporate functions. "Unit CDOs" tend to report directly to their business but have a matrix reporting relationship to the enterprise CDO. That helps prevent the development of data silos (which can lead to redundant systems and duplicate work) and ensures that best practices are shared and standards are followed. Generally speaking, unit CDOs own their respective versions of the truth, while the enterprise CDO owns the SSOT. A decentralized approach is well suited to offensive strategies because it can increase the agility and customization of data reporting and analytics. In many companies, among them Wells Fargo, CIBC, and P&G, the CDO is responsible for both analytics and data management, facilitating the ability to balance offense and defense.

It's unwise to default to a 50/50 split between offense and defense.

Finally, in choosing between a centralized and a decentralized data function, it's important to consider how funding will be determined, allocated, and spent. The budget may appear larger for a centralized function than for a decentralized one simply because it's concentrated under one

CDO. Decentralized budgets are typically more focused on offensive investments, are closer to the business users, and have more tangible ROIs, whereas centralized budgets are more often focused on minimizing risk, reducing costs, and providing better data controls and regulatory oversight—activities that are less close to business users and usually have a less-tangible ROI. Thus creating a business case to justify the latter is usually trickier. The importance of investing in data governance and control—even if the payoff is abstract—is more easily understood and accepted if a company has suffered from a major regulatory challenge, a data breach, or some other serious defense-related issue. Absent a traumatic event, enterprise CDOs should spend time educating senior executives and their teams about data-defense principles and how they create value.

CONCLUSION

Emerging technologies may enable a next generation of data-management capabilities, potentially simplifying the implementation of defensive and offensive strategies. Machine learning, for example, is already facilitating the creation of a single source of truth in many companies we studied. The promise is more-dynamic, less-costly SSOTs and MVOTs. However, no new technology will obviate an effective, well-run data-management function. Our framework will become even more relevant as distributed technology solutions—blockchain, for example—come into play.

Data was once critical to only a few back-office processes, such as payroll and accounting. Today it is central to any business, and the importance of managing it strategically is only growing. In September 2016, according to the technology conglomerate Cisco, global annual internet traffic surpassed one zettabyte (10²¹ bytes)—the equivalent, by one calculation, of 150 million years of high-definition video. It took 40 years to get to this point, but in the next four, data traffic will double. There is no avoiding the implications: Companies that have not yet built a data strategy and a strong data-management function need to catch up very fast or start planning for their exit.

A version of this article appeared in the May–June 2017 issue (pp.112–121) of *Harvard Business Review*.

Leandro DalleMule is the chief data officer at AIG.
