

Get Your Data Under Control in 5 Steps

By Scott D. Lowe

Co-Founder, ActualTech Media and
Senior IT Executive, CampusWorks, Inc.

October 2014

Introduction

Across every industry vertical and without regard for organization size, there is one fact that remains true: the data deluge is upon us. The much-ballyhooed “Big Data” phenomenon that prognosticators have warned about is here. Whether or not you’re a believer in the transformational power of data, it’s hard to ignore the fact that organizations are simply drowning in data. The reasons are many:

- For various reasons, insidious “shadow IT” has usurped the formerly centralized IT architecture, resulting in all kinds of data being stored in various places in the organization.
- Business units now have the power of using just a credit card to subscribe to new business-enhancing cloud-based services. The result: data spaghetti, multiple “versions of the truth,” and negative impact on operations.
- Central IT has become so overwhelmed with “keeping the lights on” that they simply cannot respond to end-user requests in a reasonable timeframe, forcing those business units to work on their own as best they can.
- Lack of data stewards and data architects in small and mid-sized organizations. In other words, there is no data “traffic cop” and data expansion and variance grows unchecked.
- Legacy environments are being augmented by, not replaced by, newer services. The result: constant needs to develop individualized integration mechanisms to handle synchronization of key data elements between systems.

Bad and unchecked data growth and management create major headaches for an organization and actually limit overall efficiency. In addition, without good data strategies and tools, organizations miss out on what can be incredible results, including:

- New market opportunities that can be unlocked by data itself.
- Business process improvement opportunities that can be enabled by good data.
- Improving relationships with existing customers.

In short, unmanaged data, non-integrated data, and bad data in general actively *cost a company a lot of money*. This white paper provides 5 steps that companies can take to tame the data beast and provide some potential business outcomes that can be enjoyed when data quality and consistency are firmly in hand.

Taming the Data Beast

Getting things under control takes effort, coordination, tools, and time. Here are the five steps that organizations should take to start.

1. Implement a Data Governance Structure

Patching a problem for the short term is easy. Fixing it for the long term takes effort. When it comes to the data beast, a company must take active steps to implement a solid foundation upon which to build the data future. This means taking data seriously and understanding that it's often the very lifeblood of the business. As such, for organizations that are having data issues, implementing some kind of data governance structures and processes needs to be the first step along the data journey. After all, without consistent process for handling data, any data effort that is undertaken will not necessarily align with established norms.

The very first step that needs to be taken is establishing some kind of governance structure to oversee the company's data efforts. Even in a small business, this is necessary and might be as simple as assigning overall stewardship of data to just one or maybe two people. In larger organizations, a governance structure may involve assigning a representative from each individual division to a formal team. However, these people are not there to simply represent their division. They are there to ensure cooperation between teams and to work from a high-level organizational perspective to keep data flowing responsibly around the company.

This group needs teeth; it needs to have the authority necessary to ensure that reasonable controls on data are maintained. It's useful to include a high-level executive as an equal member of the team to help make sure that data quality is considered a strategic priority.

2. Refine Software Acquisition Processes

It's become really easy for non-IT business units to acquire software and services that bypass IT. While there are some IT departments that are really slow and that hinder the business, allowing bypass of IT for critical purchases and services can have a really negative effect down the line. IT is perfectly positioned in an organization to understand broad business-wide processes and data integration needs. As such, new technology services and software acquisitions should always include input from IT on how such services are best implemented for long-term success.

Bear in mind that "success" isn't just about procurement. Success also includes the ability to leverage new tools in as seamless a way as possible and for years to come. This is where IT now needs to learn to step aside. Once the software is acquired, it's time to let individual business units do what they do best – drive their units forward. Business users who make business decisions should have a hand in developing data integration processes that meet their needs whereas in most organizations IT's expertise is in perfecting the technical aspects of the systems but not in how the data from the systems is used.

3. Understand the Full Data Lifecycle

Consider for a moment the following scenario based in higher education: a student applies for admission to a college or university. This life changing process for a student is one that is very routine for the institution and kicks off a flurry of activity, *all based on the data provided by the prospective student*. It's tempting to say things like, "Well, Admissions doesn't need a particular data element, so we're not going to capture it or keep it current." However, there is a lot of information that is critical as the student moves through an institution and more and more information builds as the student wends her way through her degree program. Eventually, the student graduates and the alumni/fundraising office assumes responsibility for the information.

There is a whole lot happening here:

- The student is initiating the process of enrolling at the college. Decisions regarding enrollment are made based on the information provided by the student.
- The student is looking for financial aid and the institution needs to stay on the right side of the law. Often, financial aid systems are separate from other systems on campus, but financial aid plays a central role in getting that student into the institution. Financial aid systems need hooks to a lot of other campus and non-campus systems, such as federal government databases. Further, even after a student is enrolled, financial aid systems must remain tightly coupled to other campus systems, such as course registration and grading systems, in order to ensure that the institution remains in compliance with a bevy of constantly changing regulations.
- During the student's time at the institution, the ERP is supposed to keep track of all kinds of information, including grades, registration, billing, and more. However, many colleges employ third party tools to meet individual operational needs. For example, to manage access to printers, colleges may implement a print management system. To help manage retention – the number of students that choose to come back year after year – many colleges turn to third party retention specialists that use their own data systems. Ultimately, much of this data needs to find its way into the ERP.
- Once a student graduates, all of the data on the student still has value. For example, students need transcripts. Further, when it comes to fundraising, efforts to raise money can be enhanced by creating affinity groups. That is, have current soccer players work to raise funds from former soccer players. The ability to leverage data to make a connection can be a powerful boost to fundraising efforts. However, at many schools, fundraising systems are separate from the rest of the ERP and must be manually integrated with one another.

In short, a whole lot happens with data from the time of application until the student no longer has an affiliation with the institution. There are many data systems that need to be kept in sync with one another. However, many schools – and beyond higher education, many companies – don't do a great job at maintaining high levels of data integrity. The results: multiple "versions of the truth" as people argue about which data is valid; lower levels of operational efficiency since data can't be reliably leveraged to automate operations; poor customer service.

It all comes down to a need to document the full lifecycle of data in the organization and understand the various ways that data is manipulated as a part of that lifecycle.

4. Implement Robust Data Management Tools

Keeping data synchronized can be next to impossible without the right tools to do the job. Sure, database experts can write custom code to handle integration needs, but how many DBAs take the time to create great documentation so that someone else can pick up the load once they leave? Further, how does an organization make sure that processes created in data management tools are executed on a regular basis to ensure ongoing data quality and consistency? Finally, where does one even find great DBAs these days? Central IT might have them, but central IT is generally swamped.

The answer: organizations need to invest in the tools necessary to enable adherence to data governance rules and that provide self-documentation so that anyone familiar with the data can manipulate it to meet organizational needs. Moreover, such tools need to be easy to use so that the user doesn't need a Ph. D. in data to use them. With the right tools, organizations can stitch together a complete data quilt from individual pieces of data fabric and create a data layer that permeates the entire organization.

Perhaps one of the most critical tools in this effort is a tool such as Astera Centerprise, a powerful yet accessible data transformation and integration product. With Centerprise, end users can easily create their own comprehensive data flows and work flows and rebuild business processes and achieve data integration efforts that once would have required teams of IT people to accomplish.

5. Initiate a Data Cleanup Project

Once data governance, product acquisition, and data tool purchases have been completed and once there is broad understanding of the full data lifecycle for a specific business, it's time to get started on righting the data ship. This involves putting into motion a broad project intended to address any data anomalies that might plague the organization. For example, are there multiple departments creating the same data elements, but doing it differently? Is the right data being captured at the right time in the organization? In the world of higher education, for example, is parent contact information being captured by Admissions so that it can be leveraged by the student life and fundraising departments? It's far easier to capture such information at application time than it is later on.

Expect data cleanup projects to take a lot of time and to be contentious. After all, in many cases, data cleanup projects require people to change their ways. Change can be tough, even when it's done for the right reasons. That makes it even more important to have good tools in place that can improve the data cleanup project.

Data-driven Business Outcomes

It's not often clear exactly what benefits might be enjoyed from undertaking what appears to be a complex project. As is the case with any other significant project, there needs to be some kind of return on investment (ROI); data cleanup projects are no exception. With that in mind, here are just a few of the positive outcomes that can come from getting the data house in order with the right tools and the right processes:

Single Version of the Truth

Good data governance and good update practices will result in having a single version of the truth when it comes to analyzing corporate data and keeping it current. Having good data practices and just one “real” view into that world streamlines decision-making, improves efficiency, and saves money.

Ability to Improve Processes

Business process improvement is a hot topic these days thanks to the many operational rewards that can be reaped. Good data, but more importantly, *consistent* data, enables data-driven process enhancements and automation that can be trusted. Now it's possible to leverage data to kick off processes. For example, when a customer order comes in, the data systems can be trusted to initiate a fulfillment process. In a college, good data means that students can be automatically billed without someone having to review every charge to make sure it's correct. The point is that while bad data can lead to bad processes, good data leads to good processes and ones that can be automated with confidence.

Easier Onboarding of New Services

With unchecked data sprawl and without an overall plan for how to handle data, new service acquisition and integration is an extreme challenge fraught with potential problems and confusion over which set of data the new service should use. With a plan and real processes and procedures for handling new and existing data elements, an organization has a much easier time integrating new services in ways that maintain the integrity of both the data originating from the new service as well as existing systems.

Summary

Quality data is the lifeblood of the modern business. As business end users become more comfortable with assuming responsibilities that would have once fallen to IT, it's critical that businesses consider the entire lifecycle of their data sets and take steps to ensure that, organization-wide, data works for the business and becomes an enabler to ongoing success.

About the Author



Scott D. Lowe is Senior Editor of EnterpriseStorageGuide.com and HyperConverged.org. Scott has been in the IT field for close to twenty years and spent ten of those years in filling the CIO role for various organizations. Scott is also a micro-analyst for Wikibon and an InformationWeek Analytics contributor. In addition, Scott has also written thousands of articles and blog postings and has regularly contributed to such sites as TechRepublic, Wikibon, and virtualizationadmin.com. Because of his unique blend of skills (CIO/strategic & Engineer/tactical) Scott is also a sought-after resource for speaking engagements, and marketing collateral, including white papers and e-books, for a variety of technology firms.

About Astera Software

Astera Software develops powerful, intuitive data integration solutions that conquer the complexities in relational and hierarchical integration processes. Whether running batch transfers or real-time integrations, Astera empowers enterprises to extract, transform, share, and migrate data through a wide range of data structures and services, on premise or in the cloud, so it can be leveraged to streamline processes, make more informed business decisions, and promote partner relationships. Astera software products are consistently acclaimed for their high performance and ease of use.