

Turning Big Data Into Useful Information



a Storage eBook

Contents...

Turning Big Data Into Useful Information

This content was originally published on the CIO Update, IT Business Edge and Enterprise Apps Today websites. Contributors: Paul Barth, Seth Earley, Loraine Larson and Susan Hall.



2 The 4 Principles of a Successful Data Strategy



6 Transforming Information Into Knowledge



9 Building a Better Ball Team With Big Data: Lessons for Executives



11 Big Data, Big Opportunities for DBAs

The 4 Principles of a Successful Data Strategy

By Paul Barth

Many Fortune 500 companies are recognizing enterprise data as a strategic business asset. Leading companies are using troves of operational data to optimize their processes, create intelligent products and delight their customers. In addition, increased demands for regulatory transparency are forcing companies to capture and maintain an audit trail of the information they use in their business decisions.

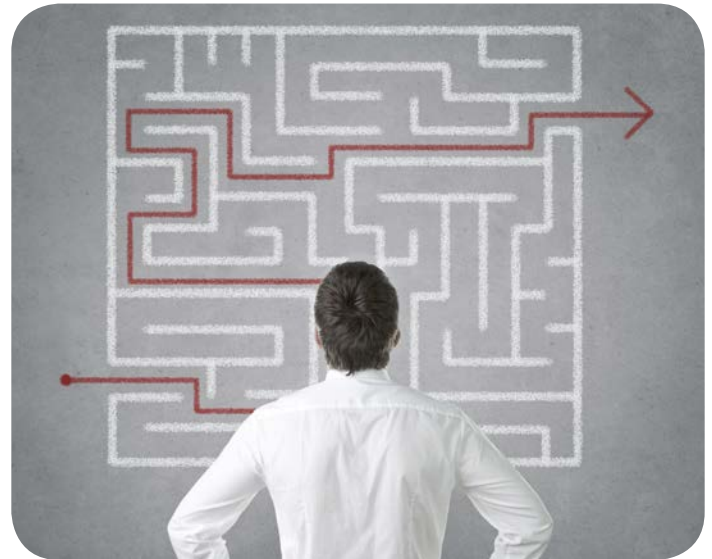
Despite this, large companies struggle to access, manage and leverage the information that they create in their day-to-day processes. The rapid growth in the number of IT systems has resulted in a complex and fragmented landscape, where potentially valuable data lays trapped in fragmented inconsistent silos of applications, databases and organizations.

IT's Not Your Fault

Experience has shown that this is not a technology problem, it is a business problem. Creating an effective data environment requires change and coordination across the board, with business and IT joined at the hip. To ensure success, they must create a practical data strategy that guides process changes as well as ongoing investments in their data assets.

In our work with Fortune 100 companies during the past 10 years, we have identified four principles behind a successful data strategy. These principles align and focus the strategy, breaking initiatives into manageable projects with a measurable business benefit. Principles are presented as questions that business and IT must answer. Those answers, in turn, help shape the framework and priorities to drive implementation of the strategy:

Question 1: How does data generate business value?
Improving the quality or accessibility of enterprise data



is not an end in and of itself. It is merely an enabler for creating business value. The data strategy must be driven by an understanding of how information can enable or improve a business process. For example, increasing cross-channel sales (a business value) requires data about your current customers and the products they own (the data); or reducing the cost of manual reconciliation for financial reporting (the business value) requires standardizing and consolidating redundant and inconsistent data across business applications (the data).

The data strategy does not need to identify all possible business benefits, but it should define several that are material to the business and measurable. Establishing some early, visible benefits is important to launching the data strategy and giving it momentum.

Question 2: What are our critical data assets? Not all data in the business is critical. In fact, most data is specific to an application, business function or transaction. Data that is critical typically has two characteristics:

- It is associated with something of long-term value to the firm (e.g., product, customer or financial information)
- It is used across multiple systems and business processes.

A high-level process flow through marketing, sales, fulfillment and finance for a top technology company would go something like this:

- Marketing creates interaction information as it reaches out to organizations and individuals through its marketing campaigns.
- As sales leads emerge, a salesperson is assigned, partners are engaged and sales opportunity information is maintained throughout the sales process.
- When an agreement is reached, terms are shared with product fulfillment to deliver the product and maintain support.
- Finance and sales validate commissions with the sales teams and partners.
- Management uses an end-to-end view of these processes to evaluate the effectiveness of its pipeline and make ongoing improvements in and across the areas.

This process analysis reveals several critical data assets and associated attributes. For example, customer organization and individual information is used by every one of the process steps. If this information is siloed and inconsistent, customers will get inconsistent messages

and service. Process owners will have difficulty measuring their effectiveness. Analyses will not reconcile. And implementing new controls or improvements will require changes within each process step.

Conversely, improvements to these critical data assets will likely yield business benefits in all five areas.

ROI

In our experience, identifying and improving critical data assets in large companies can yield tens of millions of dollars in benefit, and justify millions of dollars of investment in implementing a data strategy.

However, we believe it is just as important to keep the set of critical data assets as small as possible. Note that the most critical data asset for these subject areas is a common identifier. Maintaining the unique identity of customers, products, interactions and contracts is what links information across the enterprise. Once that is tackled, attributes can be added incrementally to the enterprise record over time.

Question 3: What is our data ecosystem? For most businesses, data is an active asset that is captured, created, enhanced and used in many business processes and applications. To manage this dynamic environment, the flows of data across systems and processes need to be organized in a coherent way.

We use a business architecture (not a technology architecture) to define core data capabilities that business and IT must create together. These capabilities organize technology platforms and business processes based on their function in the ecosystem: capturing and creating

“In our experience, identifying and improving critical data assets in large companies can yield tens of millions of dollars in benefit.”

data, cleansing and organizing it, mining business insights from it, and using those insights to drive intelligent actions in the business.

By capturing data that measure the outcomes of our actions, we create a closed loop that allows companies to use their data to test, learn and improve their processes.

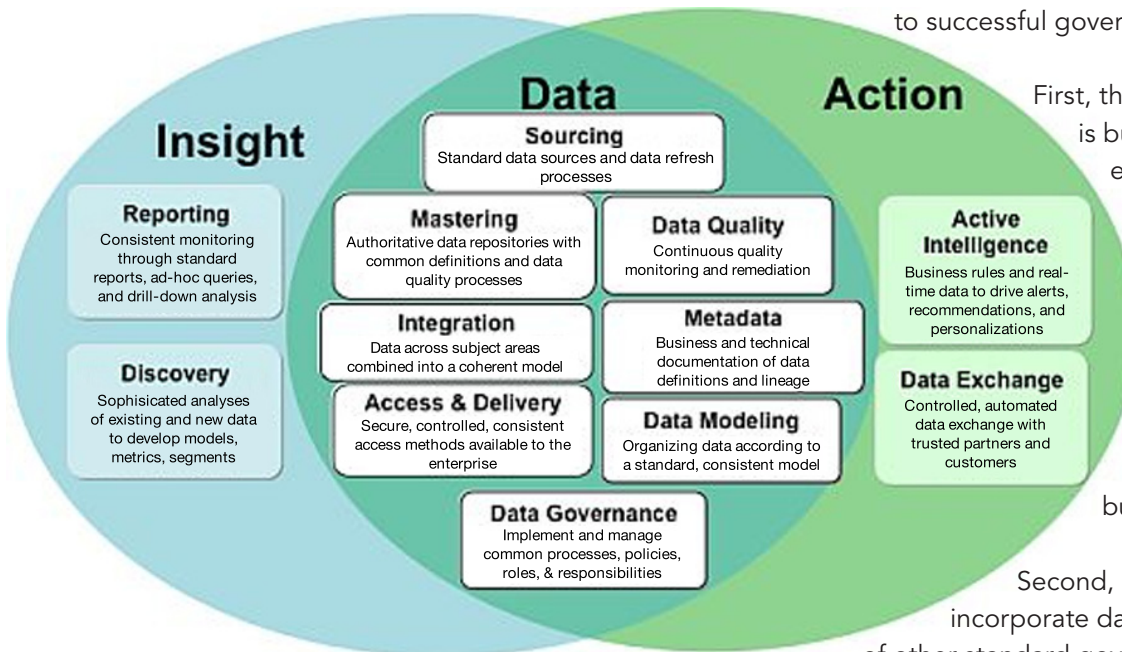
The diagram below depicts three broad classes of core capabilities: data, insight and action.

Data capabilities are responsible for creating and managing usable, high-quality enterprise information assets. These

This capabilities model can categorize thousands of applications and data repositories into 12 logical buckets that will guide their simplification and evolution toward a common strategic blueprint.

Question 4: How do we govern data? Ultimately, the implementation of a data strategy is not a project; it is an ongoing function of the company that must be governed. Because data is so ubiquitous, the governance structure must be federated, with a central governing body addressing the most important, common data and most of the data managed locally in the lines of business.

We have found several elements of this model critical to successful governance.



First, the stewardship community is business heavy, with executive business data owners supported by business data stewards who report to them. IT custodians ensure that the systems incorporate and monitor the requirements of the business.

Second, companies should incorporate data governance as a part of other standard governance procedures as much as possible, including architectural review boards, audit and risk review processes, system development methodology, and security processes. Over time, a distinct governance body for data may disappear as it is fully embedded in other business governance activities.

Third, it is important to launch data governance with a small facilitation team and some data governance related infrastructure, such as data quality, metadata and lineage tools, to provide visibility and measures to the data governance board.

include all standard data management capabilities, such as data sourcing and integration, quality and metadata management, data modeling and data governance.

Insight capabilities include tools, data and processes for management reporting and advanced analytics.

Action capabilities provision data and business intelligence to applications, business processes and business partners, and capture responses to interactions.

Conclusion

The alignment of the four principles for successful data strategy is the foundation for establishing a manageable, meaningful change in the way companies deal with data. Note that technology is not the key to success — it is merely a supporting element in the development of core capabilities.

For many firms, the first attempt at a coherent data strategy is a daunting effort, with stakeholders learning each other's language for the first time. Over time, the common understanding of how data is vital to the business establishes an effective dialogue so that truly strategic initiatives can be launched that make every business process more informed and intelligent. ■

“The alignment of the four principles for successful data strategy is the foundation for establishing a manageable, meaningful change in the way companies deal with data.”

Transforming Information Into Knowledge

By Seth Earley

Business changes faster than technology can support. That is just a fact of a CIO's life. But there are ways to improve organizational agility and better support the demands of an ever-changing enterprise.

In this article, I focus on how to create a context for amplifying the business value of information. My straight-forward proposition is that organizations that focus on core practices for effective access to and integration of information get more value from IT investments.

The resistance to addressing information access and integration head-on is that it seems like a "boiling the ocean" problem. It doesn't need to be. It is possible to focus on a business-critical solution but throughout the effort maintain an enterprise perspective designed to build information value incrementally.

Here's the recipe:

1. Focus on a specific process as a starting point.
2. Create common business language across the enterprise.
3. Embed that language in tools, systems and processes to create new business capabilities and agility.
4. Create governance and change management programs to leverage these capabilities in day-to-day work processes.
5. Apply accepted practices to unstructured content processes to promote better information hygiene.



6. Measure the information management process.
7. Measure business impact of new practices.
8. Repeat on a department by department basis.

Let's elaborate on each of these steps:

1. Focus on a specific process as a starting point — Choose a few processes that will provide significant benefits to the organization, but look at those processes in the broader context of overall enterprise organizing principles.

CIOs are frequently tasked with getting ahead of content sprawl and chaos. The problem can become widespread

and overwhelming, and it becomes difficult to know where to start.

The first consideration is to identify what processes need attention and where the right information can provide maximum impact. Although this seems like an obvious point, too often people will look across the enterprise and try to solve information management problems on an enterprise scale; making them seem intractable.

The other extreme, taking too narrow a focus, will simply lead to more siloed content and fragmented repositories. The difference is in the scope of analysis. It's important to start with specific problems to solve for specific audiences but also to zoom out, creating common structures across the enterprise.

2. Create common business language across the enterprise

— These efforts go by different names: Enterprise Architecture, Taxonomy, Information Models, Data Architecture, Information Architecture, Metadata Strategy, Master Data Management, Enterprise Vocabularies and Glossaries.

The key is to combine the business perspective with the technical perspective, and enroll the correct stakeholders in the process.

Here are some common mistakes:

- Considering this a pure technical exercise: The common wisdom is if it is data related or metadata, then it's in the IT realm. In fact, IT can own some metadata, but business users must own business language. The language used to describe solutions, markets, offerings, risks, products, features and so on is business in context, perspective and usage.
- Handing it off to library science people or information architects: Each of these roles has a valid but incomplete perspective. Avoid complete handoffs in favor of a joint exercise of business and technical groups with executive involvement and sponsorship.

- Taking too narrow or too broad a perspective: There needs to be a fine balance between focusing on a process or application and understanding the wider context. Many organizations have wasted time and frustrated participants by trying to boil the ocean. At the same time, if the focus is too narrow, there will be yet another silo of information.
- Not considering long-term governance, maintenance and change management: These projects have a fixed duration for a specific outcome, but they are never completely finished. When asked when the enterprise taxonomy will be finished, I answer, "When you are finished with sales, marketing and manufacturing — when the business ceases operations." Since everything changes in a growing evolving organization, business language will evolve over time to stay current and fresh.
- Doing it without the correct resources or approach: This area has undergone tremendous change in approach, practices and maturity over the past 10 years.

3. Embed that language in tools, systems and processes to create new business capabilities and agility

— If we look at the challenge of having common language for concepts, business terms and information objects, there are a tremendous number of opportunities for creating consistency throughout the information landscape of the organization.

This is done by ensuring that tools for accessing information evolve toward a common framework, managed through taxonomy and metadata disciplines. Representative tools include business analysis and reporting tools, search, enterprise and Web content management, portals, data warehouse, master data management, and collaborative suites.

4. Create governance and change management programs to leverage these capabilities in day to day work processes

— Governance allows for ongoing evolution of terminology and core vocabularies. Governance processes provide the leadership, resources, direction, operational guidelines and ongoing monitoring

of progress and success of the information management initiative.

The goal of good governance is to ensure the correct capabilities are being developed and deployed, sufficient resources for success are allocated, policies are being complied with and organizational change management resources are engaged throughout the program.

These efforts require the goals of the program be fully socialized, and users understand the end game and benefits as well as what is expected of them. Programs succeed or fail based on the level of socialization, buy-in and change management effectiveness.

5. Apply accepted practices to unstructured content processes to promote better information hygiene

— Most enterprises have developed poor information management habits and practices. This is understandable, as the information tsunami is growing at an astounding rate. IBM believes that in the next five years enterprise information will grow by 650 percent with 80 percent being unstructured.

The only way to avoid the increasing inefficiency and ineffectiveness that this situation will create is to begin to develop better information hygiene. This means:

- Developing and applying information lifecycles.
- Using the appropriate tools for collaboration versus reference.
- Applying appropriate resources to organize important information.
- Assigning ownership and curation across repositories.

6. Measure the information management process

— Develop metrics to measure the quality and currency of content. A big part of this is simply measuring success at getting rid of materials cluttering up the system. This can significantly improve search results, make content more usable, allow essential content to be prioritized, and ensure

that the curation process is focused on useful information.

Other useful metrics include how often critical content has been reviewed, whether tagging has been approved, the age of content, value ratings by users, heuristics for alignment with best practices, search precision and recall and, most importantly, statistical sampling of content for compliance with governance process guidelines.

7. Measure business impact of new practices — The business impact of content can be measured only through its impact on a business process.

Content supports process (e.g., knowledge base content supports customer service reps) and the performance of that function can be linked to use of high-value content: Sales processes are supported by proposal content, competitive information, marketing content, and so on; and engineering is supported by various design libraries, methodologies, test protocols and specification documents, for example.

Content quality can impact a process which in turn supports a business objective (e.g., customer retention, customer acquisition and time to market).

By creating linkage to a specific step in a process and measuring how effective or efficient that step is with supporting content versus the effectiveness with poor quality content, one can measure the impact of these new practices.

8. Repeat on department by department basis — Once this approach is initially developed and applied, create a center of excellence with the purpose of introducing these practices to other parts of the organization.

Content habits take time to develop. In fact, a baseline maturity assessment can tell you where the organization is on the learning curve and how long it will take to move through the stages that are needed to achieve excellence.

While this journey will be challenging, there really is no choice given the information avalanche that lies ahead. ■

Building a Better Ball Team With Big Data: Lessons for Executives

By Loraine Lawson

Executives are becoming a bit skeptical about Big Data's ability to deliver real business change, it seems. The problem isn't the skepticism — the problem is that it's based on misunderstanding.

Don't believe me? Good. Ask a European soccer coach or, if they're in the United States, Oakland Athletics. Or, if you prefer, ask Michael Schrage, who first noted the problem during a Big Data seminar for senior executives held in London.

Schrage is an author and research fellow at MIT's Sloan School's Center for Digital Business. He asked the room full of executives, "How much more profitable would your business be if you had, for free, access to 100 times more data about your customers?"

The response wasn't what he expected.

"... not a single executive in this IT-savvy crowd would hazard a guess," he writes in a recent [Harvard Business Review blog post](#). "One of the CEOs actually declared that the surge of new data might even lead to losses because his firm's management and business processes couldn't cost-effectively manage it."

Schrage contends this attitude is based on a misconception about becoming data-driven, but actually, it will require more human decisions.

"The reason why my London executives evinced little enthusiasm for 100X more customer data was that they couldn't envision or align it with a desirable business outcome," he writes. "Would offering 1000X or 10,000X



more data been more persuasive? Hardly. Neither the quantity nor quality of data was the issue. What matters is how — and why — vastly more data leads to vastly greater value creation. Designing and determining those links is the province of top management."

Specifically, executive leaders need to decide what matters most and make a commitment to the desired outcome — then decide how Big Data can help them get there, he says.

This advice differs from what I've heard previously, which is that [companies should approach Big Data as more of an exploration](#) to see what the data reveals, rather than a query.

As a happy coincidence, European soccer shows both can be true.

European soccer teams are super-aggressive about data. They don't monitor just the usual stats; they actually wire

up players at practice to collect data in all biochemical data, GPS data and vital signs, reports InformationWeek. Coaches can actually create player-specific drinks and nutritional supplements for each player, based on blood and saliva tests.

Here's the catch: Most of it is meaningless.

It turns out there's no correlation between how much distance a player covers on field during the game and outcome, the article notes. Likewise, traditional stats — number of tackles, shots on goal or other popular game trivia — are useless. Those measures don't impact the actual outcome of the game.

But by monitoring all this data, they did find data that correlated with winning, InformationWeek notes:

It took time for team managers and their statistics gurus to understand which of those data points to pay attention to. The percentage of completed passes compared to the number of interceptions is a good indication of consistent victory, as is an error rate lower than 18 percent.

And, in the case of Oakland Athletics, they used the data to "identify players who would be a unique asset to the team," and exploit their opponent's weaknesses, on a small budget.

The moral of that story: The teams did not know which data they needed, but they knew the outcome they wanted to create.

The same will be true for businesses. Big Data can create big outcomes, but you'll have to identify the outcome you want to identify the data that matters to your "sport" and your particular teams.

Now, it's true you shouldn't expect overnight wins. There's actually a [steep Big Data learning curve](#), as a recent CompTIA survey showed.

And that's not to say that a bit of skepticism isn't deserved — there is a lot of hype now.

"In a new era of Watson, Windows and Web 2.0 technologies, any organization that treats access to 100X more customer data as more a burden than a breakthrough has something wrong with it," Schrage warns.

So don't dismiss Big Data as useless, just because it's being hyped. Instead, Schrage suggests CIOs and executives:

- Rethink how your organization adds value.
- Treat Big Data and the algorithms that run them like managing top talent.
- Ask "What value matters most, and what marriage of data and algorithms get us there?" instead of how do we get more value from more data? ■

"The moral of that story: The teams did not know which data they needed, but they knew the outcome they wanted to create."

Big Data, Big Opportunities for DBAs

By Susan Hall

Amid all the talk about Big Data, questions keep popping up about how the database administrator (DBA) fits in. The role is changing — and it varies by company, according to Bert Scalzo, database domain expert for Quest Software.

“One of the biggest concerns I keep hearing is how the DBA role has either lost importance or gained importance. There’s been a change, and there’s disagreement about what the change is,” he said.

DBAs More or Less Important?

Scalzo said some people feel the DBA role is becoming more important, due to the large amounts of data companies are accumulating — and storing, thanks to inexpensive disk space and cloud storage. That doesn’t mean there are more DBAs, however.

“It used to be that you could only have so many terabytes or petabytes before it went to secondary storage. Now that’s kind of old-fashioned. Now you can keep data live until there’s absolutely no reason to have it,” he said. “In that case, the DBA is doing more. It’s not that the number of DBAs is decreasing, but the amount of data that companies are keeping is increasing exponentially with the same amount of resources.”

Those people see DBAs growing more important, but at the same time more overwhelmed, he said.

Some other companies treat data as a commodity and may think they don’t even need a DBA on staff. Such companies are more likely to outsource DBA work.

“If it’s in the cloud, Amazon’s making sure it performs well, doing the tuning, optimization, backup and



recovery and all that other stuff. That’s part of the service,” he said.

Broader Definition of a DBA

Traditional DBA work still needs to be done, said Neil Raden, CEO and principal analyst at Hired Brains Research.

“If you use a broader definition of a DBA as not so much an administrator, but a data management person, those skills are going to be very valuable,” Raden said, noting that data management skills encompass a broad range of knowledge in areas such as data quality and data integration.

He said some of the most exciting job opportunities in Big Data involve statisticians and analysts who can work

with predictive modeling and machine learning. Yet DBAs tend not to be mathematicians.

Raden said there are two kinds of DBAs: designers and instance tuners. Designers “look more closely at building and modifying schema, they work with ETL people and how they get introduced to different databases,” while instance tuners “are mostly involved with writing SQL query, developing and maintaining indexes and other types of physical optimization.”

“So one is more deep down in the weeds of how the machine works, and the other is more up in the air with the data,” Raden said. “The people with the data are more likely to find, at least partially, their skills useful in this different environment. The other guys have hitched their wagon to Oracle’s [latest version] or DB2. Those databases aren’t going anywhere,” he said.

And there are plenty of job opportunities for them. The unemployment rate for database administrators stood at a mere 2.8 percent during the first quarter of 2013, according to the U.S. Bureau of Labor Statistics.

Open to a New Paradigm

Especially with cloud services, specific skills with Linux, Windows, HP-UX or AIX become less important, Scalzo said.

“The biggest thing [in making the transition] is not to presuppose or apply your Golden Rules that you’ve learned over your last 10 or 20 years,” he said. “Good database design in a relational environment might not be good design in a column-oriented data store. How

you define data structures relationally might be different than how you define them in a noSQL database. It’s kind of an unlearning/learning process that you have to embrace. You still apply common sense; you just use your experience to gauge that common sense in a new paradigm.”

An open mind and inquisitive capabilities are the two most important qualities for DBAs in this new and changing data environment, he said. “You’ve got to have an open mind and be hungry to learn and see how what you used to know works in a new environment.”

‘Immense Opportunity’

In a blog post, Gwen Shapira, senior database administrator at data infrastructure management company Pythian, argues that [DBAs are prime candidates for a new role](#) called Hadoop cluster administrator. (Citing other better technologies, Raden isn’t convinced DBAs need to run out and learn [Hadoop](#).)

Shapira writes that any organization using a Hadoop production cluster larger than 20 to 30 nodes will require a full-time administrator. That job will be similar to traditional DBA work: responsibility for the performance and availability of the cluster, the data it contains and the jobs that run there.

While that admin will need to be able to write some small jobs in Java, for Hadoop code, alternatives such as Streaming, Hive and Pig will do. Hadoop admins will need a deep knowledge of Linux, however.

“So one is more deep down in the weeds of how the machine works, and the other is more up in the air with the data.”

As for how traditional DBAs should be positioning their careers, Shapira said in an interview with Enterprise Apps Today that they can opt to continue what they have been doing, since relational databases are not going away any time soon.

She believes there is “immense opportunity” for DBAs.

“In the age of Big Data, data professionals are more visible to the organization than ever, and they are seen as the critical resource they always were,” she said. “A few years ago, executives only noticed the database when it stopped working. Now they are looking at the database and they think of all the information it holds as an opportunity: How can we get more value out of the

information we have here? What other information can we collect to support our decision making? How can we use data to enhance our products?”

She offered some advice for DBAs who want to use this opportunity to take their careers to the next level: Step out of the server room and into management offices.

“Talk to business unit managers to find their pain points. Which information do they need? What are the gaps in the current systems? Managers are happier to talk to technical people now than they’ve ever been,” she said. “Show them how your skills can help them get more value out of data, because this is what Big Data means to them.” ■