Designing for data quality

BY Todd Shutts, Robb Montgomery AND Hans Kiamzon

When properly managed, data is a valuable asset that drives informed decision-making throughout your organization. Understanding the data life cycle and implementing effective data maintenance are essential to accessing the high-quality data you need, when you need it.

All around us, the data landscape is evolving. Ninety percent of the world's data was created in the last two years. Every day, according to *Forbes*, another 2.5 quintillion bytes of data are created. And 200 billion devices will be connected to the internet by 2020. All kinds of businesses are experiencing an explosion in the amount and types of data available, as well as in the number of devices that are feeding data to the enterprise.

Data should be viewed as an asset, just like your product or customers. In fact, the data on an asset can be more valuable than the asset itself. That's because quality data helps businesses make decisions that impact both the top and bottom line. Yet, few companies have consistent access to data that is valuable at the point of decision-making.

Many organizations assume if they have data, they can do effective analysis. But to be valuable, data needs to be carefully managed from creation to consumption. This requires a thorough understanding of the data life cycle and, in particular, how data quality is maintained. It also requires careful planning to make sure you have the data you need — not only to solve existing problems, but to address new issues as they come up.

Realizing the value of your data

Having consistent access to high-quality data allows you to conduct rigorous business analysis. This in turn allows you to extract patterns and facts that will inform your decision-making.

For example, a utility that has invested millions in a turbine needs to monitor that asset's performance to make sure preventative maintenance is done on schedule. This requires accurate data on how many times the turbine rotates, whether its temperature is consistent within a given range and other factors. Having access to this data allows the utility to spend thousands of dollars on maintenance rather than millions to replace the entire turbine.

Data can also be used to prioritize repairs or customer service. For instance, a power loss impacting thousands of suburban customers might look like a complicated repair. So, you might decide to complete other, smaller repairs first. If data analysis indicates the issue can be resolved with one truck, it's smarter to address the big problem before making repairs that affect fewer customers.

In both examples, high-quality data underpins essential business analysis. In other situations, using data to optimize maintenance or repairs may not be warranted. It may make more sense to simply replace a low-cost piece of equipment when it fails or complete minor repairs on the spot.

Understanding the data life cycle

Businesses tend to think of data as a static entity that can be plugged into all kinds of analytics. But, it's important to understand that data is subjected to many processes and activities as it moves through its life cycle.

Five categories of data — transactional, master, reference, metadata and reporting — exist in various sources throughout any organization. In a retail organization, for example, transaction data is generated by a customer-facing sales app when a transaction is performed. The data is then transmitted to an operational data store or data warehouse. When needed, the data is exported to a reporting tool for use in research and analytics. At the end of its useful life, it is retired or archived.

Avoiding Common Problems

Progressing from creating data to supporting analytics is one of the biggest challenges businesses face today. A few common problems are usually to blame:

- **Access.** Businesses want to make decisions in real time, like reordering popular products as they sell out. Yet, they may only have access to their inventory data. Real-time decision-making can't happen if data is loaded once a day or once a month from a back-end system.

- **Context.** To make decisions, you need to know what your data represents, but data can have different meanings across the enterprise. Plus, data collected for one purpose often isn't suitable for other analytics. This typically happens when data is considered in isolation, rather than in the context of a system or application.

- **Data quality issues.** Low-quality data may be malformed, duplicative or missing. Poor data quality can often be traced to errors made during data entry, lack of business rules defining acceptable data values and problems associated with migrating data from one system to another. Sometimes data is missing simply because it wasn't collected.

To fully realize the value of your data, it's essential to address these data quality issues in advance.

Managing your data

Effective data management starts with a data governance framework. Essentially, this is a set of rules defining the types of data you will collect and how that data will be represented. More precisely, the Data Governance Institute defines data governance as "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods."

Enterprise data architecture is evolving, wholly as a result of the pressure that technology is imposing and the growing desire for self-service analytics. Traditionally, data governance implementation involves working with a functionally and technically experienced data management team to execute the following activities.

The first step in creating a data governance framework is to identify the functional use cases that define your business. A typical retail use case is a customer interaction at the point of sale, while a utility use case might be online bill payment.

Defining use cases helps you think in terms of real-life scenarios. You can then deconstruct to define major data entities, their relationships to other entities and their attributes. In a retail business, major data entities include products, customers, stores, warehouses, suppliers and vendors. A utility company's data entities include customers, field workers, power stations, substations and meters.

Once you've established your major data entities, you need to define acceptable representations of that data. A retail customer might be defined as someone who makes a purchase at a store or online and has provided an email address, street address and phone number. A utility power station might be defined in terms of monthly billing with key data entities such as customer, service point, usage and rates.

With this level of detail, you can start to understand how to define data quality. For example, every phone number needs to be 10 digits, or every email must include the "@" symbol with a domain name. Finally, all departments must agree to represent vendors, materials or other shared data in the same way. Once you have defined a set of data rules, those rules can be implemented in an IT system to maintain the quality of your data throughout its life.

It was accepted that these implementations would involve lengthy delivery timelines before providing the end user with the high-quality data, tools and artifacts initially requested. With the volume and sources of data growing dramatically, the velocity at which this data is being delivered — and the growing desire for self-service analytics from end users — means the traditional implementation approach is no longer suitable for meeting the needs of an organization's decision-makers.

The enterprise's data architecture must evolve by developing the speed and agility needed to meet these information demands. An experienced team will recognize a set of characteristics of those organizations that have adapted to meet these data management challenges:

- **Collaboration.** Fostering collaboration among the business, technology and leadership groups helps to promote data being a shared asset, develop a common vocabulary and determine how to optimize the managing of the data through its life cycle.

- **Promotion of self-service capabilities.** Getting data into the hands of the decision-makers as expeditiously as possible is realized through providing self-service analytics. By empowering end users with this capability, there is less reliance on project delivery timelines, allowing the enterprise to react, anticipate and plan.

- **Focus on capturing metadata.** Metadata is a key component to enabling analytics on data. Metadata allows an organization to catalog its data, empowering various types of analyses.

- **Evolution from historical reporting to predictive and prescriptive analytics.** As an organization is able to better understand, manage and derive insights from its data, it is able to unlock higher value-added reporting capabilities.

- **Adoption of lightweight toolsets/platforms.** Developing speed and agility from a technical perspective involves adopting tools that can easily integrate with the organization's various applications and data sources with minimal infrastructure footprint to extract the metadata necessary for analysis.

Accessing data on demand

The best way to make sure you can access the quality data you need, when you need it, is to work with a data management team that has both functional and technical knowledge and experience. Functional knowledge identifies use cases and then define key data entities and attributes, while technical know-how helps you implement appropriate tools and solutions.

An experienced team of consultants will have a thorough understanding of the data life cycle and data quality issues, which they can use to anticipate and prevent problems. The team will also provide insight into how your data program should be designed and delivered in order to support the analytics that are crucial to your business.

Just because you have data doesn't mean you have quality data. Working with a data team that combines functional and technical knowledge and experience will see that you have access to high-quality data across your enterprise and throughout the data life cycle.

Biographies

Todd Shutts, director of Business Intelligence & Analytics at 1898 & Co., a business, technology and security solutions consultancy, part of Burns & McDonnell, is a software builder and entrepreneur who leads teams in creating viable, innovative technology that delivers value to our customers. Throughout his career, Todd has been involved in the defense, power generation, financial technology and retail industries. This diverse background allows him to work with teams to assemble technology strategies that help our clients build successful products, processes and organizations.

Robb Montgomery is a director of Enterprise Systems & Integration at 1898 & Co., part of Burns & McDonnell. where he leads the Enterprise Systems business line specializing in digital transformation. His team focuses its efforts on cloud solutions, "internet of things," data governance, integration solutions and a variety of other services that assist our clients' success. With more than 25 years of experience across all areas of IT, Robb has knowledge and experience in complex transformation program management and project delivery utilizing an array of methodologies.

Hans Kiamzon is a data architect at 1898 & Co., part of Burns & McDonnell. Since 2007 he has worked in systems development and integration, primarily delivering management solutions for electric utilities, retail, networking, financial services and other industries. Hans has firsthand experience with a variety of data integration platforms, data management tools and data conversions. Currently, he is focusing on enhancing the company's data integration and governance capabilities.

About 1898 & Co.

1898 & Co. is a business, technology and security solutions consultancy where experience and foresight come together to unlock lasting advancements. We innovate today to fuel your future growth, catalyzing insights that drive smarter decisions, improve performance and maximize value. As part of Burns & McDonnell, we draw on more than 120 years of deep and broad experience in complex industries as we envision and enable the future for our clients. For more information, visit **1898andCo.com**.