aws

The data-driven enterprise

By Mark Schwartz, Enterprise Strategist, AWS

Introduction

There is a lot of talk these days about the data-driven enterprise and the need to become one. But what exactly does it take to become data-driven, and why is it so important in today's digital environment? What practical steps can an enterprise take to make data fundamental to its mindset and practices? And what is the connection between data and that other priority of the digital age—business and technical agility? This eBook will show what it means to be data-driven and give some examples of how companies are using data to drive their businesses. We'll also connect the dots between becoming data-driven and agility, digital transformation, and continuous innovation.

Data-driven organizations strive to ***base their strategic business decisions*** on the evidence provided by data—which requires a certain rigor and, at the same time, an ability to innovate based on identifying—within the data—opportunities that can lead to new products or markets. They also come to treat data as an asset they can use both to improve customer interactions and to increase efficiency. In other words, they ***analyze*** data to inform decision-making and ***use*** data to serve their customers. Data can be the basis, for example, for personalization, dynamic pricing, market expansion, product innovation, or supply chain optimization.

But until recently, enterprises found it difficult to use data in these ways because they thought of data solely in the context of transactions; as a result, they locked it away in siloed databases that were excellent for transaction processing but less suited to open-ended analysis. Our mental model was that of the invoice or the order form: "Please give me 20 widgets at a price of $100 per widget." Or, "Please pay me for 20 widgets at $100 per widget." Data was performative and imperative—a stimulus or artifact of conducting a transaction. Today, the value of data goes far beyond its transactional role.

How can we think about this value in financial terms, and how can we maximize it?

The business value of data

Each piece of data can be used in any number of analyses that will drive business results. It has value, then, in making possible the results that are obtained from those analyses. For example, if the enterprise analyzes its historical transactions and, as a result, finds ways to optimize its supply chain, thereby reducing costs, then the data has played a role in enabling that cost reduction. Consequently, data has a business value that stems from its potential use in increasing profits or accomplishing mission objectives.

It is easy to find instances of data being used for its non-transactional value. Johnson & Johnson, for example, uses the transactional data it has stored in the cloud to improve physician compliance, optimize its supply chain, and discover new drugs. Nike collects data on customer achievements to drive the customer's digital experience in NikePlus. Lyft collects and stores the GPS coordinates of all of its rides; when they analyzed it, they found that 90% of rides overlapped with other rides from nearby locations. This insight led to the creation of Lyft Line, a service that allows passengers to share a car and receive discounts of up to 50 percent.[1]

Because these uses can lead to future profits—even if the profits are not yet being realized—we can think of data as a financial asset (although in most cases a non-GAAP asset). It is no surprise, then, that the data a company has accumulated can be a factor in the acquisition value of the company or may enable it to form partnerships with other business ventures. Witness, for example, Microsoft's acquisition of LinkedIn, with its data on 433 million customers, for $26.2B or the bankruptcy proceedings of Caesars Entertainment Operating Corp. Inc. in 2015–2017, where creditors argued that the data on the 45 million customers in its Total Rewards customer loyalty program was worth $1B and was its most valuable asset.[2]

It is helpful to think of data as having business value as a kind of financial call option—that is, it gives us the opportunity to make changes in the supply chain or launch a new product but does not obligate us to do so. We can exercise the option or not, depending on how valuable the data indicates that the new business will be. It is here that we have had trouble finding the value of the data asset: Valuing a call option is considerably more complicated than calculating the ROI of a projected stream of cash flows. As a result, enterprises often neglect the value; but as I show in my book ***War and Peace and IT***[3], many of the techniques of agile IT delivery result in this kind of option value.

[1] AWS case studies. See https://www.youtube.com/watch?v=6A1tOFqvgek, https://aws.amazon.com/products/databases/, and https://aws.amazon.com/solutions/case-studies/lyft/.

[2] Both examples from https://sloanreview.mit.edu/article/whats-your-data-worth/. A detailed analysis of the Caesars bankruptcy can be found at https://turn-around.org/sites/default/files/11.%20Paper%20-Caesars.pdf. The bankruptcy was exceedingly complex and the value of Total Rewards was included with other assets, so it is not clear what value was ultimately attached to it.

[3] Mark Schwartz, *War and Peace and IT: Business Leadership, Technology, and Success in the Digital Age* (Portland, OR: IT Revolution Press, 2019).

Data and agility

Value is created not just by the data per se but also by the tools and processes we have in place to analyze it and produce those business outcomes. In today's digital world, fraught with rapid change, uncertainty, and complexity—disruption, you might say—we need to use data to support business agility and to respond quickly and flexibly to changing circumstances. Agility is what enables organizations to turn rapid change into opportunity and to avoid disruption by responding nimbly to competitive threats. Enterprises in the digital age have learned that they need to get early versions of products to market quickly and evolve them through continuous feedback from the market.[4]

The last few years have brought techniques for building agility into the product development ***process***, including, for example, Agile software development, DevOps, and Lean software development. The cloud has been used to speed up the delivery of IT capabilities, for both software and hardware. Team-based organizational structures have made it possible to mobilize the resources to meet changing needs. All of these developments have helped enterprises make their processes more agile.

But agile processes are only one part of the story: The company's data itself must also be agile. It must be easily available for uses that are unexpected and constantly changing. It must be accessible and meaningful. Employees must have tools easily available to work with the data and the skills to do so. It is this ability to use data flexibly—to make it available for new uses that we don't know about in advance—that is the missing link in achieving enterprise agility and distinguishes the agile organization from one that has merely adopted the frameworks and trappings of agile models. Business agility requires data agility. A data-driven enterprise is a master of both.

[4] Ries, Eric, *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses* (New York: Crown Business, 2011).

This focus on bringing agility to data is new. As long as data was only transactional, we could lock it away in highly structured databases whose structure reflected the way it would be used for those transactions. Our tools were relational database systems such as Oracle or SQL Server, whose strengths are in transactional processing. We used the data to conduct the transactions themselves and to produce operational reports to support the transactions.

To the extent that we paid attention to privacy, we enforced it by strictly limiting access to the data rather than searching for ways to make it available within the bounds of privacy guardrails. Instead of "privacy by design," we practiced a sort of "privacy by obscurity."

Yes, there were attempts to free data for ad-hoc analysis with so-called business intelligence (BI) systems. But the tools have now advanced far beyond what BI systems were meant to do: We now have machine learning, a range of purpose-built databases to handle different types of data, algorithms for massively parallel processing, vast amounts of unstructured data like video and speech, IoT devices that deliver streams of sensor-derived data, and…well, just vast amounts of data. With these tools, we can free our data from its transactional and operational context.

More importantly, we have realized that being data-driven is not just a technical challenge but also an organizational one. **To be data-driven, an organization must think differently about how it makes business decisions and how it interacts with customers. It is a commitment to the value of data, a kind of organizational humility that says, "the data knows better than we do."**

How can we make our data available to be used in unexpected ways; that is, how can we use it flexibly to give us business agility? How can we apply it to bring rigor and creativity to business decision-making? How can we change business culture to take advantage of this new flexibility?

And how can we put appropriate control guardrails around the data to safeguard its privacy while at the same time allowing it to be used flexibly and quickly?

There are really two questions:

1. How can we bring agility to our data?
2. How can we use data to bring agility to our business?

Agility for data

How can we bring agility to our data?

To achieve business agility, we'll need to be poised to respond to unexpected changes in the business and competitive environments, and we'll need to create innovations that are truly novel—and so, we will need to be able to put our data to work in ways that we don't necessarily anticipate when we collect it.

Our challenges:

- Our data is probably locked away in transactional, relational databases and probably siloed in ways that make it inaccessible to different parts of our organization.
- We may not have the right analytical tools, or they may not be available to the right people at the right times.
- Our models for security and privacy are ad hoc, as we perhaps never contemplated using the data for exploration. Most likely, we are fostering privacy simply by making the data as inaccessible as possible.

Our goals:

- Maximize the data's availability, subject to guardrails for privacy and confidentiality.
- Foster transparency across the enterprise by breaking down information silos.
- Offer employees the appropriate tools to explore the data in unplanned ways and in ways that take advantage of the latest advances in analytics.
- And be sure to have the expertise to interpret the data, both rigorously and creatively.

In "Analytics without Limits: FINRA's Scalable and Secure Big Data Architecture," John Brady, the CISO of the Financial Industry Regulatory Authority (FINRA), frames these objectives elegantly by saying that he wants to ***lower the cost of curiosity***. He refers to cost in its widest sense, including the time it takes to draw inferences from the data and the risk in making it available. FINRA's business is to explore the 37 billion or more transactions that take place in the financial markets every day, looking for patterns of fraud. Since they don't always know in advance what a pattern of fraud looks like, they must rely on the expertise of their analysts to spot suspicious behavior. Their task is all about curiosity: They want their analysts to examine data with inquisitiveness as to what patterns appear and why. The task of their IT organization is to reduce the cost of that curiosity and the effort that an analyst has to exert to explore a hunch.

Brady's idea applies across organizations and roles. Can a marketer easily explore data to find unexpected patterns in consumer purchasing activity? Can operations explore data to identify performance optimizations or to diagnose problems in operating processes? Can finance explore data to concoct new ways to drive performance or to slice and dice data to drive executive decision-making? Can IT leaders test their hypotheses about how to optimize cloud spending with rigor and creativity?

Curiosity drives innovation and improvement. Agile data allows employees to freely explore ideas, hunches, hypotheses, and conjectures at the speed of thought and to promote new ideas with the data to support them.

To make data agile, an enterprise needs to address how and what data it gets, how it preserves that data, how and under what conditions it makes the data available, and what tools and skills it has for working with that data.

1 Get the data

To use the data nimbly, we must first have the data. And given the unknown uses to which we will put it, we need to collect ***more*** data than we know how to use. That, in a nutshell, is what "big data" is about. Fortunately, with the cloud, the cost of storing data is low and declining. We can, therefore, instrument our business processes to produce data, lots of it, and make it available for analysis. For example, the Internet of Things (IoT) applications often include sensors that blast a stream of data points into the cloud that the enterprise can analyze immediately or store away for future analysis. Enterprises can also now work with a much wider range of data types: video, text, and speech, for example. The possibilities for using all of this information in novel and interesting ways is tremendous.

GE Oil and Gas, for example, pulls an MRI-like device they call a "pig" through their oil pipelines to collect over 750 TB of information that helps them spot potential problems in the pipeline infrastructure. Hudl has collected about 10 PB of video and other data that sports coaches can review with players. Peloton gathers data from their exercise cycles and analyzes it to provide insights to their customers. And Airbnb accumulates about 50 GB of data each day for fast analysis in the cloud, using Amazon Elastic MapReduce (EMR), a tool that allows large volumes of data to be analyzed quickly in parallel.[5]

5 https://aws.amazon.com/solutions/case-studies/ge-oil-gas/,
https://aws.amazon.com/solutions/case-studies/hudl/,
https://aws.amazon.com/solutions/case-studies/Peloton/,
https://aws.amazon.com/solutions/case-studies/airbnb/.

2 Store it

Once we acquire the data, we must store it to make it available for analysis. Traditionally, we stored data in a structured format based on our expectations about how it would be used transactionally. For example, we might have a field in a database for "quantity ordered" and another field for "unit price." We would collect the data to fill these fields and file them away by slotting them into the appropriate blanks in the database, knowing that we could always multiply those to values to derive a total price. By forcing the data into such a mold, we made it useful for transactions, but we might have lost information that could have been useful for analysis. This was the relational database model.

The past few decades have been dominated by the use of these relational databases, which are very well suited to efficient processing of old-world volumes of transactional data in ways that are known in advance ("multiply unit price by order quantity"). But when you are working with non-transactional data or operating at tremendous internet scales of transactions or managing data that does not slot easily into pre-defined "data fields," there are now much better alternatives, purpose-designed for the cloud.

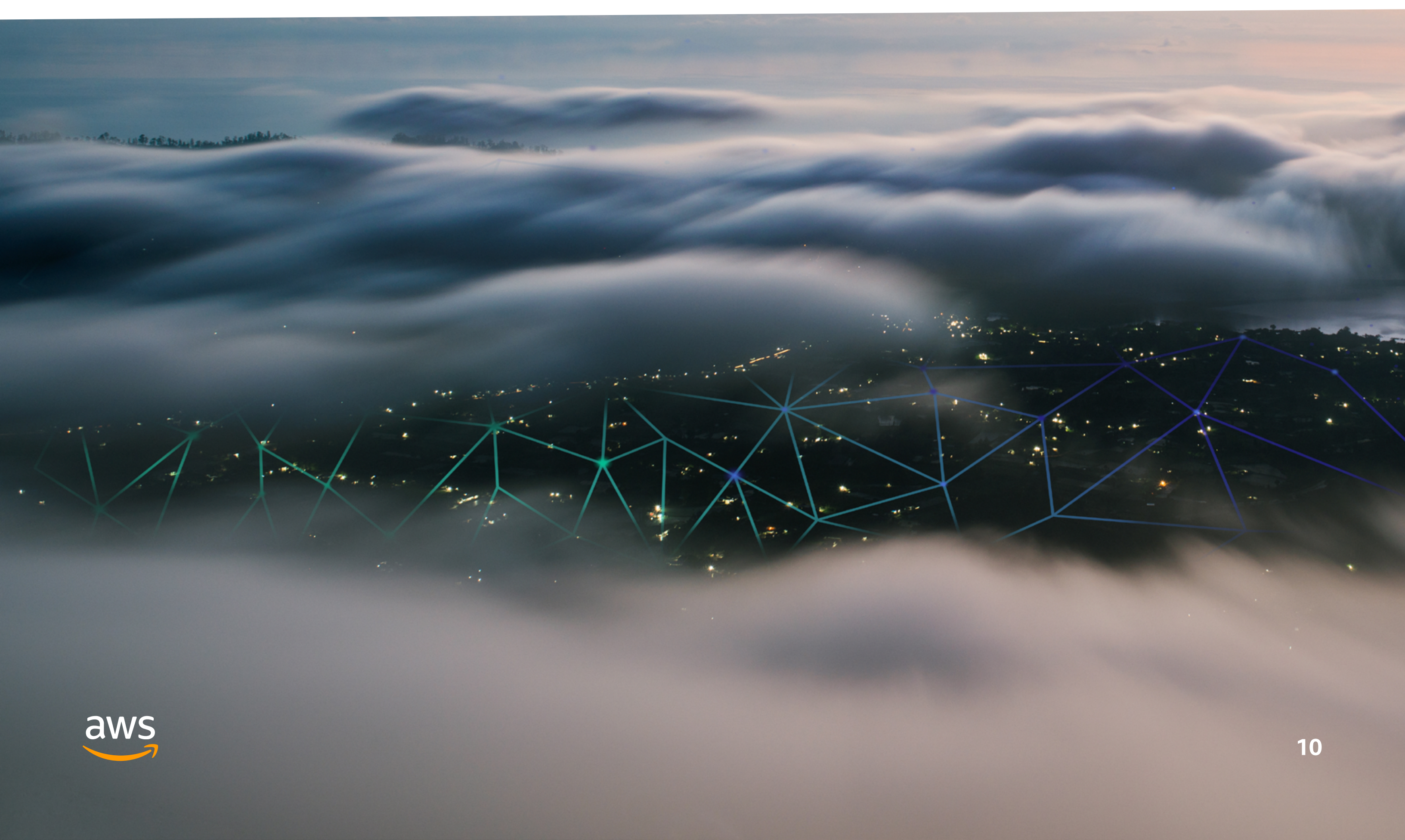
For example, Amazon Timestream is a database designed specifically to manage time-series data (like the data produced over time by an industrial sensor or by tracking market activity over time); Amazon Quantum Ledger Database is intended for the type of data used in blockchain (data whose history must be verifiable, using techniques like cryptography); and Amazon Neptune is designed for representing complex connections and relationships, like social networks. Enterprises are no longer limited to what they can force-fit into a relational model.

Better still (for agility), data that will be used for yet-undetermined analysis can be stored in a flexible repository called a data lake, where each piece of data is stored simply in the form in which it was received. The power of the data lake lies in the tools that can be used to analyze it: tools that let you combine heterogeneous information, mixing together structured and unstructured data, data from different organizational silos, and data in large quantities. Today's tools can apply machine learning algorithms and statistical analyses, and they can work with natural language text, video, and speech.

In other words, the data lake meets the enterprise need for storing data before it knows all the ways it will be used. We can pour data into the lake from different business silos and analyze it all together. We can quickly set up a way to pour data from a newly acquired company into the lake and thereby gain transparency into its operations, and we can integrate its data with our own. The magic that makes this all possible is: (1) the low cost of storage, (2) the availability of tools that work with loosely structured, heterogeneous data, and (3) the availability of services that lets you push data into the data lake at high bandwidth and asynchronously (just send the data toward the data warehouse as you receive it, and it will get there as quickly as it can, no need to wait—sort of like an email).

3 Make it available

The next step in bringing agility to data is to make it available—when and where it is useful. (Note that I didn't say when and where it is ***needed***. I'm talking about agility and innovation here.) The model that is often used today is one of self-service provisioning. When an analyst is curious, he or she can spin up a set of tools and a subset of the data to analyze without having to request and wait for someone else to provide it. The resulting freedom lets the analyst pursue a train of thought, a "flow," rather than proceeding in a stop-start way that destroys creativity—or, you could say, that increases the cost of curiosity. The cloud is an important enabler for this, as it allows new work environments to be provisioned, used, and then discarded when no longer needed. It also makes it easy to put guardrails in place to protect privacy (more on this below).



4 Provide tools

A data-driven enterprise makes the appropriate analytic tools available to its employees easily and quickly, often through a self-provisioning model, as described above. A wide variety of software and services is available: If you want to perform traditionally structured queries against the data, for example, you can set up a data warehouse based on the data in the data lake, or you can provision a tool that lets you do old-school, SQL-type queries directly against the data lake.

But today, there are many more possibilities. You can, for example, visualize your data with modeling tools, and you can construct scenarios and ascertain their consequences. Today's analytics revolution is all about artificial intelligence and machine learning, which opens up new possibilities for what we can do with our data: predict outcomes, spot anomalies, categorize data, analyze sentiment, discover patterns, guide robots…and much more.

For example, Capital One is using machine learning to detect fraud while still maintaining high levels of customer service. T-Mobile uses machine learning to improve its customer service by having it predict what articles will be most helpful to the customer and making them quickly available to customer service agents. Sky News, in their coverage of Britain's royal wedding, used AWS machine learning to recognize the faces of celebrities in the crowd and identify them for the TV audience. And Formula 1, Major League Baseball, and the National Football League are all using machine learning to enhance the viewer's experience of their sports.[6]

To apply machine learning, you train a model based on earlier data sets and then apply it to new data as it is observed. In AWS, there are three general approaches to machine learning: (1) use a pre-trained model such as Amazon Rekognition, which has already been trained to recognize objects in images, or Amazon Lex, which has been trained to understand intentions expressed in natural language, (2) train and apply your own model based on any one of the common algorithms used for machine learning, using Amazon SageMaker, or (3) use your own algorithms and training approaches, if you have employees skilled in machine learning, by working directly with Amazon infrastructure that is optimized for machine learning.

With tools such as these, enterprises can unleash the creativity of their employees and find new ways to put data to use.

[6] https://aws.amazon.com/machine-learning/customers/innovators/capital_one/,
https://aws.amazon.com/machine-learning/customers/innovators/t_mobile/,
https://aws.amazon.com/blogs/media/sky-something-new-at-the-royal-wedding/,
https://aws.amazon.com/machine-learning/customers/.

5 Upskill

The next important element in extracting value from your data is to make sure you have employees with the right skills…in addition to a sense of curiosity. This is why data scientists are in such high demand today. Yes, there are plenty of tools available for even people with little skill or experience in statistics. But to really make the most of data, and to do so with rigor, it is important to have people with a good understanding of how to make correct inferences from data.

For a simple example, those of us with less statistical experience tend to over-rely on averages, even when looking at an entire distribution of values can often lead to important insights. In one case I remember from my time as CIO at USCIS, we were looking to reduce the time it took us to process certain types of applications. We created dashboards to track the average amount of processing time, but each change we tried seemed to have only a small impact on the metric. What we had missed was that the small number of applications that raised national security or fraud concerns took much longer to process, thereby skewing the average. We had no way to control how long those took. Although our improvements applied to the great majority of cases, because of the highly skewed average, we couldn't really see their impact. When we realized the problem and began monitoring, say, the 85th percentile completion time, we could identify the significant impact our changes had on the vast majority of cases. We had the data, the tools, and the access…we had just lacked the skills to draw the correct inferences.

Data-driven decisions can also be poorly founded when the data is presented (even unintentionally) in a misleading way. In his book ***The Visual Display of Quantitative Information,*** Edward Tufte shows how data can be distorted or obscured by the way it is presented.[7] Again, an enterprise that wants to be rigorous in its use of data must ensure that it has the right skills in analysis and presentation, as well as the data.

[7] Tufte, Edward R., *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 2015).

6 Provide guardrails

Before we can make data available for novel uses—to satisfy curiosity, so to speak—we must put guardrails around it for privacy and confidentiality. Data-driven enterprises practice "privacy by design," deliberately establishing safeguards based on planning and foresight. They gain speed and flexibility down the road by making sure that they have already considered what needs protection and have set up automated ways to do so. In fact, the recent European Community General Data Protection Regulation (GDPR) requires privacy by design.

The cloud provides many tools for setting up automated access controls and does so at a granular level that lets you give employees access to precisely the data they should have access to. There are ways to track the provenance and validity of the data, to encrypt or obscure it, and to restrict access on a field-by-field basis or record-by-record basis. In other words, you can specify which customers' data an employee has access to and which pieces of data associated with those customers the employee can view. Amazon Macie even uses machine learning to identify which data in your data lake is personally identifiable information (PII) and track how it is used. Or you can choose to manage data only at an aggregated level or with information masked or anonymized. The flexibility is there; each data-driven enterprise must make responsible decisions about privacy given the type of data they steward.

Many other challenges arise in using the vast amounts of data that the enterprise has available. It is often a challenge to accurately connect data from different IT systems pertaining to a single individual, especially in countries like the US that do not have a single national ID system. Data can be inaccurate not only because of mistakes made in data entry but also because of limitations in the IT systems that collect the data. For example, there are IT systems that only allow for a surname and a given name, which imposes inaccuracy for people who have more than two names.[8]

Regardless, the goal of a data-driven enterprise is to make data available to drive rigorous and accurate decision-making and continuous innovation. It requires collecting and storing data for flexible use later, making it and the right tools available without friction to those who will use them, ensuring privacy and confidentiality by design, cultivating the skills to make valid inferences, and solving the data hygiene problems that can lead to poorly informed decisions. This is what it means to bring agility to data.

[8] For some great stories about IT systems insensitive to real-world scenarios, consult Gojko Adzic's Humans vs Computers.

How can we use data to bring agility to our business?

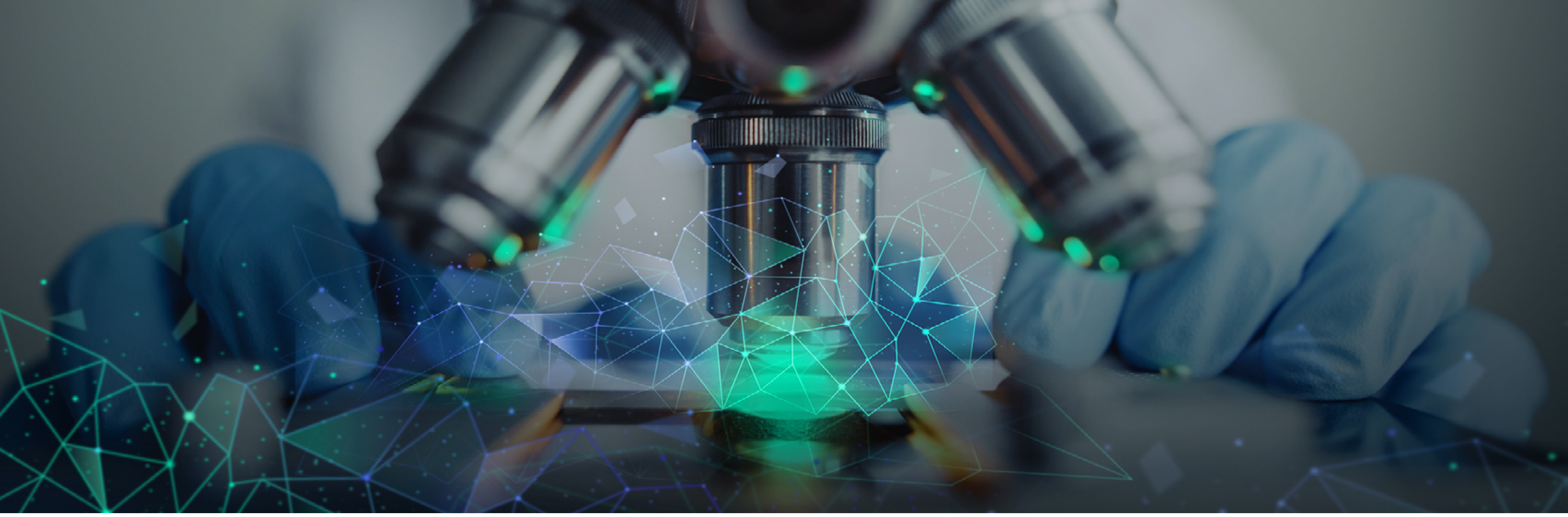
An agile business in the digital age proceeds by trying an idea, getting feedback, and then adjusting course—and doing so repeatedly. This fast-feedback approach lets the company innovate (at low risk, high speed, and low cost) and reduce investment risk by testing ideas before committing to them. It results in a good fit between the company's products and the markets they are intended to serve and ensures that the company is solving the right problem in the right way at the right time.

Fast feedback

Feedback, in this sense, does not mean asking customers whether they like a new feature or product. More commonly, data-driven enterprises use quantitative feedback—the kind of feedback that is gathered by watching how customers ***actually act***—or by monitoring changes in market behavior or other metrics.

For example, companies often improve the usability of their websites through A/B testing; that is, by trying two variations on a piece of the design (usually one variation is the current, status quo version, and the other is a new piece of design they are considering introducing). They show some customers version A and some version B. They collect data on the customers' activity and analyze it in relation to the outcomes they care about. If they want to decide whether to make a button green or red in order to maximize the number of times it is clicked, then they can show some users a green version and some a red one, and see which gets more clicks. Expedia and Netflix are examples of companies that routinely do A/B testing, drawing on large amounts of data from a data warehouse in the cloud.[9]

[9] https://www.youtube.com/watch?v=k8PTetgYzLA.



The powerful approach of learning and adjusting through feedback goes far beyond just A/B user interface testing. New product ideas, for example, can be tested by creating a "minimum viable product," the smallest and simplest version of the product the company can use to gather information on whether the product will be successful or what needs to be changed to make it so. Marketing strategies, promotions, technology alternatives—all of these can be tested through trial and measurement to reduce uncertainty. And the key to doing so is gathering data and making it available for analysis.

The technique of using minimum viable products and fast feedback is described in Eric Ries' book ***The Lean Startup***.[10] According to Ries, at any given moment, a startup holds two hypotheses: a value hypothesis, about how their proposed product will create value for customers, and a growth hypothesis, about how the company will be able to grow its market—that is, get customers to use the product. The minimum viable product is the smallest product that will give the startup information to confirm or refute these hypotheses, at which point it can make changes and re-test them with the market.

This set of practices does not just apply to startups or to new product development. It has become central to the way organizations, including large enterprises, achieve business agility by changing course based on their learnings. If an enterprise is thinking of developing a new IT system for use by its own employees, it presumably has a hypothesis about how that IT system will deliver the business outcomes that are proposed in its business case. That hypothesis should be tested, and changes should be made based on what the data shows.

As a result, agile practice requires data: To learn and adapt, the enterprise has to collect data on the impact of its new initiatives and use it to inform those initiatives. Agility further requires that the enterprise sense changes in its business environment, so it can respond appropriately to maximize its business outcomes. A data-driven enterprise not only brings agility to its data but also uses data to support its agility.

[10] Ries, Eric, *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses* (New York: Crown Business, 2011).

Culture and process change

Becoming data-driven, in this sense, requires a very different way of making decisions; it is a deep cultural change for many organizations. In the past, we might have made decisions by crafting detailed plans, analyzing options with the available data, and choosing the option that—given only the available data—appears to deliver the best outcomes. In the digital world, we refuse to accept only the data that is available at the instant the plan is created. Instead, we design experiments to yield additional data and then incorporate that data into our decision-making. We resolve uncertainty by generating new data.

An example is the technique for IT governance that we devised at USCIS. Instead of writing a large requirements document and handing it over to the technologists for implementation, we simply handed over a business objective. In one case, for example, we noticed that a skilled case processor (a “status verifier”) could process about 70 cases a day, and our business objective was to make that number much higher. In another business case, we found that a number of paper files got lost in transit as we moved them between processing locations, and we wanted to eliminate those losses.

For each of these objectives, we began by creating a dashboard that showed the key metric: the number of cases per day or the number of files that were missing. Instead of writing a requirements document, we created a cross-functional team of business operators and IT technologists, and we charged them with improving the metric. We gave them the tools to make changes to IT systems and business processes quickly and then monitored the dashboards with them. They tried small, incremental changes and monitored the results every day. Based on what they saw in the data, they could decide what to do next to maximize the outcome. And management could decide whether to continue funding the initiative or direct the funds elsewhere. The result was a data-driven, reduced-risk, lightweight governance process that delivered value quickly.

This leads to another important point: Accountability is enhanced by transparency. By making data widely available, we made the team’s progress visible. As a result, oversight bodies could constantly revisit the investment decision, either investing more or less, redefining objectives, or stopping the investment entirely. Results were the only gauge of success, and results could be achieved quickly. But those results had to be supported by the data.

Spotting patterns

Another area where data can promote agility is through sensing changes or recognizing patterns in the environment. For example, machine learning can be used to detect and respond to anomalies. We can train a machine learning model with historical or routine data so it becomes used to what is "normal" and then apply it to find activity that is not normal. This technique can be used, for example, to spot fraudulent transactions or network intrusions by hackers. Or to spot equipment on a factory production line that is diverging from its normal behavior and might have to be repaired or replaced—and to do so before it actually fails.

When we collect large amounts of data, we may find that we can identify relationships that we didn't know were there. Social media companies build large databases of relationships between people. Homeland Security might find that a potential terrorist they are investigating once lived at the same address as someone who is already known to be a terrorist—which might lead them to ask questions when they next encounter the person. A number of fraudulent immigration applications might turn out to have all been prepared by the same immigration lawyer. Here, we have moved well beyond simply using data to process transactions: We can now find important and interesting relationships between those transactions. But once again, we don't know exactly what relationships we might find; agility, flexibility, and curiosity are the keys to deriving value from data.

To cite one more example of using data to "keep an eye on events," the existence of a data point can serve as confirmation that an activity took place—for example, when audit trail logs are created automatically. By following the trail of activities, auditors may be able to validate compliance or investigate improper activity. Blockchain is often used to store data that confirms that activities took place—for example, a transfer of money between two parties or an approval of a contract by the parties involved. By using automated guardrails and audit data to establish compliance, enterprises can often avoid heavyweight compliance processes that reduce agility.

There are, of course, challenges in using data to support business agility. As we noted above, it requires skill to draw the appropriate inferences from data. The data does not always tell us what action to take: We have to interpret it and make good decisions. Often, we face a trade-off between false positives and false negatives—for instance, if we use the data to spot anomalous transactions to identify potential fraud, we run the risk of flagging too many transactions as anomalous and annoying our customers or flagging too few and allowing fraud to sneak through. The larger the data set becomes, the more likely that meaningless patterns will emerge or that important patterns will become buried in the sheer number of potential connections. Noise accumulates along with signal.

In closing

A data-driven organization is one that puts data to work to improve business outcomes, both by using data to drive a rigorous decision process and by making the data available for stimulating innovation and providing value to customers. When data is locked into an inflexible framework, siloed, or difficult to get at, it becomes a barrier to business agility, preventing the company from responding to opportunities or from getting products to market quickly. Even worse, when a business ***doesn't*** drive its processes and investments through the use of data, it is foregoing important contact with the market it is trying to serve or passing up feedback that could help it succeed better in its initiatives. A data-driven organization, on the other hand, uses data to gain agility and uses agility to make its data more valuable.

About the Author

Mark Schwartz is an enterprise strategist at Amazon Web Services and the author of ***The Art of Business Value, A Seat at the Table: IT Leadership in the Age of Agility, and War and Peace and IT: Business Leadership, Technology, and Success in the Digital Age. Before joining AWS, he was the CIO of the US Citizenship and Immigration Service*** (part of the Department of Homeland Security), CIO of Intrax, and CEO of Auctiva. He has an MBA from Wharton, a BS in Computer Science from Yale, and an MA in Philosophy from Yale.

Mark Schwartz,
Enterprise Strategist, AWS

Read more from Mark Schwartz »