MIT TECHNOLOGY REVIEW CUSTOM

Produced in partnership with

ORACLE

The Rise of Data Capital

"For most companies, their data is their single biggest asset. Many CEOs in the Fortune 500 don't fully appreciate this fact."

– Andrew W. Lo, Director, MIT Laboratory for Financial Engineering

"Computing hardware used to be a capital asset, while data wasn't thought of as an asset in the same way. Now, hardware is becoming a service people buy in real time, and the lasting asset is the data."

– Erik Brynjolfsson, Director, MIT Initiative on the Digital Economy

Executive Summary

Data is now a form of capital, on the same level as financial capital in terms of generating new digital products and services. This development has implications for every company's competitive strategy, as well as for the computing architecture that supports it.

Contrary to conventional wisdom, data is not an abundant resource. Instead, it is composed of a huge variety of scarce, often unique, pieces of captured information. Just as retailers can't enter new markets without the necessary financing, they can't create new pricing algorithms without the data to feed them. In nearly all industries, companies are in a race to create unique stocks of data capital—and ways of using it—before their rivals outmaneuver them. Firms that have yet to see data as a raw material are at risk.

The vast diversity of data captured and the decisions and actions that use that data require a new computing architecture that includes three key characteristics: data equality, liquidity, and security. The pursuit of these characteristics drives the reinvention of enterprise computing into a set of services that are easier to buy and use. Some will be delivered over the Internet as public cloud services. Some corporate data centers will be reconfigured as private clouds. Both must work together.

New capabilities based on this new architecture, such as data-driven tailoring of products and services, will yield not only radical improvements in operational effectiveness, but also new sources of competitive advantage.

Data Capital Creates New Value

In 2013, an international ring of cybercriminals attacked one of the world's largest online ticket marketplaces. But they didn't do anything as obvious as hacking in and stealing credit card information. Instead, they hijacked legitimate customer accounts using stolen log-in credentials. Then they purchased tickets to sporting and music events using the customers' payment methods, resold those tickets, and pocketed the cash.

That kind of online theft is big business. In 2015, identity fraud affected 13.1 million people in the United States alone at a cost of $15 billion, according to Javelin Strategy & Research. Digital thieves are often hard to spot because they hide in plain sight, masquerading as legitimate customers while carrying out fraudulent transactions.

At the time it was breached, the ticket marketplace had nearly 40 million accounts. How did the company figure out which ones were compromised? Its data scientists analyzed eight years of transaction and customer histories to identify the anomalies that indicated this scam.

But fraud detection is a cat-and-mouse game: As soon as the bad guys know someone is on to them, they try a new angle. Algorithmic detection is the only way to keep up, and training on mountains of data is the only way for algorithms to learn the difference between good behavior and bad. Since the 2013 attack, improved algorithmic policing at the time of purchase has reduced fraud attempts at that ticket marketplace by 95 percent.

In this example, it's easy to focus on the algorithm as the hero of the story. But it's really just an engine; data is the fuel that makes it run. Without the data, the fraud-detection capability cannot exist. Data is as necessary to the marketplace as its ticket inventory, the people who manage the business, and the money that pays for it all.

This simple idea has big implications. It means that data is now a kind of capital, on par with financial and human capital in creating new digital products and services. Enterprises need to pay special attention to data capital, because it's the source of much of the added value in the world economy. "More and more important assets in the economy are composed of bits instead of atoms," notes Erik Brynjolfsson, director of the MIT Initiative on the Digital Economy.

In fact, many companies that are light on physical assets but heavy on data assets—for instance, Airbnb, Facebook, and Netflix—have changed the terms of competition in their respective industries. While many incumbent companies possess comparably large troves of data, they don't exploit it nearly as well. These companies must adopt a new mindset, Brynjolfsson says: "They should start thinking of data as an asset."

A 2011 study conducted by Brynjolfsson and colleagues at MIT and the University of Pennsylvania supports the concept of data as a capital asset. Based on surveys of

> "Some financial services companies still don't seem to understand that they're sitting on a gold mine, and that if they ignore it, the gold mine can just turn into a trash heap. They're literally throwing away pearls of wisdom because nobody is looking at the data."
>
> – Andrew W. Lo, Director, MIT Laboratory for Financial Engineering

nearly 180 large public companies, researchers concluded that businesses that emphasize "data-driven decision making" (DDD) performed highest in terms of output and productivity—typically "5 to 6 percent higher than what would be expected, given their other investments and information technology usage," said the report. "Collectively, our results suggest that DDD capabilities can be modeled as intangible assets which are valued by investors and which increase output and profitability."

> "Common observations about the abundance of data are misleading. Instead, the issue is one of variety—the fact that there are enormous amounts of scarce, or even unique, pieces of data."
>
> – Paul Sonderegger, Big Data Strategist, Oracle

In fact,"for most companies, their data is their single biggest asset," notes financial economist Andrew W. Lo, who is Charles E. and Susan T. Harris Professor of Finance at the MIT Sloan School of Management and director of the MIT Laboratory for Financial Engineering. But, he notes: "Many CEOs in the Fortune 500 don't fully appreciate this fact." Perceptions about the value of data vary widely from industry to industry, says Lo, who is also a principal investigator at the MIT Computer Science and Artificial Intelligence Laboratory. "Uber, Amazon, eBay—these companies really understand predictive analytics. Big-box retail stores also understand the value of their data." But many other industries have yet to focus on data as an asset.

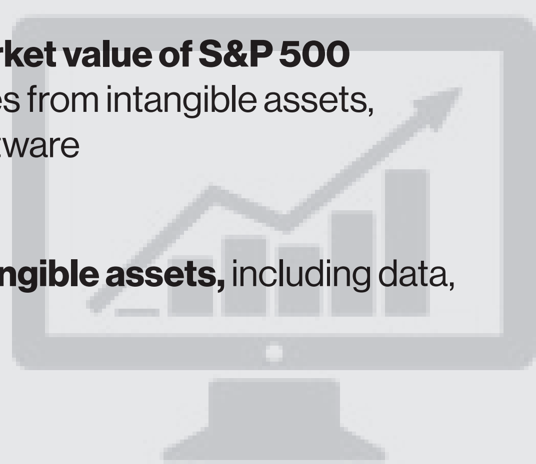
For example, Lo says, "some financial services companies still don't seem to understand that they're sitting on a gold mine, and that if they ignore it, the gold mine can just turn into a trash heap." Specifically, some financial-services institutions gather, but don't save, valuable demographic data about their clients and their activities, Lo says: "They're literally throwing away pearls of wisdom because nobody is looking at the data, and because it's taking up space."

Data capital encompasses all digital data: truck movements captured by GPS trackers; "likes" and shares recorded by social media; purchases, returns, and reorders held in enterprise transactional systems. "The challenge is embracing this diversity and figuring out how to use it at scale," says Paul Sonderegger, Oracle's big data strategist. In addition, Sonderegger says, data capital requires new computing infrastructure and deep understanding of how to create applications that analyze and use the information.

Data's Economic Identity

To call data a kind of capital isn't metaphorical. It's literal. In economics, capital is a produced good, as opposed to a natural resource, that is necessary for the production of another good or service. Data capital is the recorded information necessary to produce a good or service. And it can have long-term value just as physical assets, such as buildings and equipment, do. "With data capital, if you know something about your customer or production process, it might be something that yields value over the years," says Brynjolfsson, who is also the Schussel Family Professor at the MIT Sloan School of Management.

Early Attempts to Value Data Capital



84% Percentage of the **market value of S&P 500 companies** that comes from intangible assets, including data and software

$8 trillion Possible **value of intangible assets,** including data, in the United States

Sources: Ocean Tomo LLC; 2015 Intangible Asset Market Value Study; The Wall Street Journal

Brynjolfsson acknowledges that the concept of an intangible asset can be challenging to grasp. "You can't see data the way you can see buildings, and people are inevitably biased against things that they can't see," he says. "It's a blind spot. But this is something that is more and more important to the world economy. It's not visible, but it's still something that smart managers have to keep an eye on."

However, data capital plays by its own rules. "It shares characteristics with several other kinds of capital, but combines them into a unique mix found nowhere else," Sonderegger notes. Ultimately, he says, data is a non-rivalrous, non-fungible experience good:

> The explosively fast growth in digital services raises the specter of disruption before an incumbent even knows it's in trouble. And with the increase of digitization and datafication in every industry, every company has exposure to data-capital disruption. The question is what to do about it.

• **Non-rivalrous.** Capital equipment, such as a truck, for example, can be used by only one person at a time. Economists call this kind of resource "rivalrous." It's the same with financial capital. You can invest a dollar in only one opportunity at a time. However, data is different. "A single piece of data can fuel multiple algorithms, analytics, and applications simultaneously," Sonderegger says.

• **Non-fungible.** True commodities, such as barrels of oil, are fungible, or substitutable. For instance, you can substitute one barrel for another. But a given piece of data, such as a price, can't be substituted for another, such as a consumer sentiment score, because each carries different information. "As a result, common observations about the abundance of data are misleading," Sonderegger says. "Instead, the issue is one of variety—the fact that there are enormous numbers of scarce, or even unique, pieces of data."

• **An experience good.** Experience goods, such as movies or books, are things whose value is realized only after you've experienced them. With data, this means you only know its value after you've put it to use, Sonderegger says. This is why choosing which movies or books—or data—in which to invest your time, money, and attention always carries with it a degree of uncertainty.

If this all sounds too academic, beware. Seeing data as capital is both highly practical and urgent in real-world business, because it reveals the playbook of potential digital disruptors.

Data Capital Disrupts

Data capital is one of the most important assets of every online consumer service created in the past decade. Google, Amazon, Netflix, and Uber have all realized that data is more than just a record of something that happened. Data is raw material for creating new kinds of value, especially digital services. And sometimes, these digital services—whether offered on their own or wrapped around physical products—disrupt incumbents and reorganize entire industries.

The concept of disruptive innovation initially was made famous by Clayton M. Christensen, a best-selling business author and Kim B. Clark Professor of Business Administration at Harvard Business School. Christensen pointed out that some technological innovations allow new entrants to change the terms of competition in established markets, threatening incumbent market leaders many times larger.

In some cases, the changes reverberate well beyond the original markets. For example, Lo says, "It may not be obvious to a bricks-and-mortar businessman what Uber actually offers, but there's little doubt that it has had a massive impact on transportation, fuel economy, and people's lives." But when looking at data, Lo says, many CFOs don't focus on the business and economic implications. "They tend to view data as just a technology issue, not a strategic asset."

New technologies can enable two kinds of disruptive strategies. The first, *low-end disruption*, lets new entrants compete directly with incumbents, but at a dramatically lower price. Amazon used that strategy successfully in the retail-book industry; WhatsApp did so in messaging, pushing the price to zero. The second type of strategy, *new-market disruption*, allows companies to target non-consumption, reaching customers who previously lacked the time, skill, or money to access a product or service. For example, Netflix's original DVD-by-mail service was so much more convenient than going to a store that it put the leading brick-and-mortar DVD rental companies out of business.

One key characteristic of disruptive innovations is that while, by traditional measures, they may not be as good as existing products, they compensate by being better in other dimensions, such as price, convenience, or new features. This advantage becomes disruptive when the new offering not only improves in its new dimensions of performance, but also becomes good enough in the old ones that the traditional market embraces it and leaves incumbent offerings behind.

But if disruptive strategies are so well known, then why are they still threatening? How can they still take incumbents by surprise? In a word, speed. Some of Christensen's original examples, such as the steel minimills that marginalized integrated mills, took decades to unfold. By contrast, Uber grew to more than a $50 billion valuation in six short years. Digital services grow at blistering speeds because of network effects, low distribution costs, and data-driven learning curves. But their value, Brynjolfsson says, "can vanish very quickly, and it can scale and be replicated very quickly as well. It creates a situation of volatility and winner-take-all markets."

Network effects happen when each user of a solution benefits from the fact that many peers have chosen the same solution. The telephone system is a good example, as is any social-media service. Again, network effects tend toward a winner-take-all outcome in a given market.

The costs for companies to distribute digital services are low because Internet and wireless providers have invested heavily to build out their networks, and 2.6 billion users worldwide had smartphones in their pockets at the end of 2014, according to Ericsson's 2015 Mobility Report. The Swedish technology giant's research predicts that 6.1 billion people—or about 70 percent of the world's population—will use smartphones by 2020. The ability to immediately find and access information online reduces consumers' search costs to learn about a new service and figure out whether it's worth their time to try.

> Data capital will extend the reach of algorithmic decision-making. As managers incorporate more data into their thinking, more kinds of decisions will reveal their inner logic. Choices once thought to be strictly the domain of human beings will become the province of machines.

The data produced by consumers' interactions with these services fuels data-driven hypotheses about new features to add, then generates data about the adoption of those capabilities. Amazon did that with its recommendation algorithms, and Facebook did it by inserting news into members' feeds.

This all adds up to explosively fast growth in digital services, raising the specter of disruption before an incumbent even knows it's in trouble. And with the increase of digitization and datafication in every industry, every company has exposure to data-capital disruption. The question is what to do about it.

Three Principles of Data Capital and How to Apply Them

Competitive strategy means creating unique value in a unique way, economist Michael Porter, Bishop William Lawrence University Professor at Harvard Business School, has said. It's not enough to provide products or services that your customers can only get from your company. Your company also has to create those offerings in a way your rivals can't easily copy.

In his classic 1996 *Harvard Business Review* article "What Is Strategy?," Porter described this hard-to-copy way of creating value as a company's "activity system." Activities are the processes a company carries out every day—the way it runs marketing campaigns, designs products, bundles offerings, provides support, manages risk, and protects patents. Every activity uses a combination of financial, skill-related, technology, information, or process resources.

Data's Economic Identity

Data is:	In contrast:
Non-rivalrous. A single piece of data can be used by multiple algorithms, analytics, or applications at once.	A single dollar or piece of capital equipment can be used by only one party at a time.
Non-fungible. One piece of data can't be substituted for another, because each carries different information.	With a true commodity, such as barrels of oil, one unit can be substituted for another.
An experience good. The value of information can only be attained by knowing the information itself. But once known, the information can be easily replicated.	The value of a durable good can only be attained by possession of the physical item, not simply knowing about it.

Because information is the only resource both used and produced by every activity in a company, the digitization and datafication of more and more daily activities has a big impact on competitive strategy. The good news for incumbents: The tools of data capital are available to all companies, not just to startups. In fact, enterprises have a distinct advantage in amassing stocks of data capital because of the volume of their interactions with customers, suppliers, and partners. The three following principles—which Lo, of MIT Sloan, describes as good guidelines for data capital—show how to exploit this advantage.

Principle #1: Data Comes from Activity

From a data-production perspective, activities are like lands waiting to be discovered. Whoever gets there first and holds them gets their resources—in this case, their data riches. But not all that glitters is data gold; some activities are more valuable than others.

The rise of data capital demands a new enterprise computing architecture. Strengthening a single tier of the IT stack will just create new bottlenecks someplace else. The entire stack must beef up.

All activities produce information, but they don't produce digital data unless they involve an application, device, or sensor. Companies that have been able to see and pursue this foregone data—the information rising off activities, places, and things like so much evaporating steam—have profited greatly from it. When Google deployed fleets of cars onto the world's roads to capture imagery, distances, and wireless network IDs, and to associate all that information with GPS coordinates, few understood that the cars were amassing data capital that could be used to create search, navigation, and ad-placement services.

Utilities installing smart meters, brokerages creating mobile advisory apps, travel sites recording all the offers visitors *don't* click on—all of these are colonizing new data lands.

It's difficult to know which activities will yield the most valuable data. The answer will vary from industry to industry and company to company. Naturally, a company should focus on activities that reinforce its competitive advantage, the things that make it unique. However, to make educated guesses, a company should look first to its biggest revenue and cost drivers, especially where it interacts with the outside world. Interactions with customers, suppliers, and partners are particularly crucial because rivals are probably looking at them, too. It's imperative to digitize key activities before the competition does. The reason: If you're not party to an activity when it happens, your chance to capture its data is lost forever.

For example, the Australian supermarket chain Coles is experimenting with a palm-sized kitchen device for making grocery lists. Scan a barcode or just tell it you want milk, and it adds the item to an online list. Through the device, Coles can gather data not just about the items customers want, but also about how and when customers make their shopping lists, opening new possibilities for targeted ads and improved service.

Digitizing activities means involving sensors or mobile apps in the activities in some way. *Datafying* activities means expanding the observations you capture about them. "Datafication"—a term introduced by Kenneth Cukier and Viktor Mayer-Schönberger in *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Eamon Dolan/Mariner Books, 2014)—runs contrary to data-management orthodoxy, which tries to settle on the minimum dataset necessary.

For instance, an airplane manufacturer captures tens of millions of data points from every test flight of its latest passenger plane. Engineers use this data to speed up the delivery of safe planes, but what else will they do with all of those observations of such a variety of flight characteristics? Not even the manufacturer knows yet. But the option value on potential uses of that data is likely greater than the cost to capture, store, and experiment with it.

Principle #2: Data Tends to Make More Data

Algorithms that drive pricing, ad targeting, inventory management, or fraud detection all produce data about their own performance that can improve their future performance. This data-capital flywheel effect creates a competitive advantage that's very hard for other organizations to overcome, which makes experimenting with data from new digital touch points doubly important.

For example, Uber uses a dynamic pricing algorithm when demand for rides is exceptionally high. Surge pricing, as it's called, raises the normal prices of fares to provide an incentive for more drivers to get on the road, bringing supply in line with demand. But the price can't rise so high that customers leave the service in disgust. The surge-pricing multiple, which has risen as high as eight times the normal fare, is determined algorithmically based on Uber's data about supply and demand in different cities under different conditions. Because the algorithm runs on proprietary data, only Uber can use it. More important: Every time surge pricing runs, Uber captures more data about price sensitivity under yet another set of conditions, fine-tuning the algorithm's future performance.

> A data-centric approach to security still incorporates traditional tactics like hardening the perimeter. But it also recognizes the realities of data capital. Data doesn't sit still; it moves around. It gets used in different systems, by different people both inside and outside the enterprise.

The challenge for most companies is to figure out where this data-capital flywheel effect will be most powerful. While careful thinking by domain experts is useful on this front, real-world experimentation is the true key.

Businesses must develop both infrastructure and new skills in experiment design, data analysis, and interpretation. Managing that process will require new tactics to handle conflicting, data-driven perspectives on the same topic.

Principle #3: Platforms Tend to Win

Network effects, so important in the growth of digital services, come in two types. *Direct network effects*—which affect users in a single market, like people on a particular social network—are a double-edged sword: They can shift into reverse as users hop to the next thing. Consider the collapse of the social network MySpace after Facebook's rapid rise.

Indirect network effects happen when there are two or more markets, and a user in one benefits from choosing the same solution as the majority of the other group. People who use credit cards, for example, want to carry a card that more merchants will take, and merchants want to take the kind of card that more people have. The same goes for application developers and computer users in their choices of operating systems, game developers and gamers in their choice of consoles, and potential employers and job-seekers in their choice of job-market sites.

In each of these examples, there is a platform between the two markets that reduces the transaction costs for the two sides. Growth on one side attracts users on the other, and vice versa, leading to a winner-take-all or winner-take-most outcome. Platforms increase their staying power by adding new features that support more linked and adjacent activities for players in each market.

Digitization and datafication of more activities brings platform competition to the doorstep of companies that may have never seen it before: health care providers and payers, manufacturers, property developers, maintenance and repair providers. For example, drones now generate near-infrared images of crops that indicate nitrogen uptake from the soil. This information can be fed to a spreader to target the right mix and amount of fertilizer for different areas of the field. In addition, data on crop yields from seed makers and on weather-related risk from insurers could improve output even further. The maker of the tractor in the middle suddenly finds itself competing to be the platform for agricultural services.

Companies must reevaluate their strategic positioning by looking for platform threats, Sonderegger says. "Ask where a new entrant might set up as a platform between two markets in the industry," he advises. "Then consider how your company might digitize and datafy activities key to that potential platform before the would-be disruptor gets there."

The Three Principles of Data Capital

1. **Data comes from activity.** Digitize and datafy key value-creating activities before rivals do.

2. **Data tends to make more data.** Data-driven algorithms create data about their own performance that can improve future performance, creating a virtuous circle.

3. **Platforms tend to win.** Digitization and datafication of more daily activities bring platform competition to more real-economy industries.

Data Capital and Enterprise Architecture

The rise of data capital demands a whole new enterprise-computing architecture. Strengthening a single tier of the IT stack, such as analytical tools or data management, will just create new bottlenecks someplace else. The entire computing stack must beef up. Simultaneously, the advent of cloud computing, buoyed by continuous price-performance improvement in processing, memory, storage, and bandwidth, has kicked off a renaissance in enterprise-computing capability.

Large enterprises are reinventing their corporate computing environments as a new set of services delivered by cloud infrastructure. They buy some from public-cloud vendors, while they build others in corporate data centers that have been converted to private clouds.

That new computing environment figures into the data-capital equation as well, says Brynjolfsson. "Computing hardware used to be a capital asset, while data was not thought of as an asset in the same way," he says. "Now, hardware is becoming a service people buy in real time, and the lasting asset is the data. People are increasingly thinking about their intellectual-property issues, thinking about data rather than computer servers."

The hybrid cloud environment is the basis for data-capital computing. The reconfigured data management, integration, analytics, and application capabilities of data-capital computing adhere to three critical principles: data equality, liquidity, and security and governance.

> "Software is being embedded in more and more products, from autos to television to prescription medications. They're all generating data. That creates opportunities to analyze data, and that data analysis creates value."
>
> – Erik Brynjolfsson, Director, MIT Initiative on the Digital Economy

Data Equality

Data has a shape. For the past 40 years, the most common shape has been the rectangular table of the relational database. Its familiar row-and-column structure has been a powerful way to capture selected observations about the real world, with broad applicability and economical use of computing resources. But this reductionist approach to creating data is now being supplanted by an expansionist approach. We don't want just selected observations of the world; we want the whole thing.

Some of these observations still do lend themselves to relational rectangles, but their schemas (the columns for each row) may be different, because they capture the attributes of different kinds of customers, products, processes, or transactions. This increased variety of carefully designed structures expands the diversity of data capital. In addition, the capture of more data from more activities brings data in a riot of new shapes into the stock of data capital. Pictures, audio files, sensor logs, geographic maps, network relationships, and other kinds of information all create data that doesn't lend itself to rows and columns.

Data equality means capturing and keeping data in its original shape and format. A new generation of data management software—some of it open-source, some from commercial vendors—departs from the relational model, more easily accommodating a greater variety of data shapes.

For example, Hadoop, one of the most popular open-source projects of the past few years, isn't a database at all. It's a file system, like the one that stores documents and spreadsheets on your laptop. As such, it doesn't require the design of a data model before storing data. Instead, developers just pour in files, images, or database extracts as they are.

NoSQL databases, which get their name from using "not only SQL" to query data, are based on a decades-old idea called a key-value store. Unlike a relational database, which requires each record to conform to a common set of attributes, a NoSQL database lets each record have its own set of attributes. This is useful, for example, when storing data from sensors in aircraft: Each device detects a different set of atmospheric conditions or forces, but analyzing them together is essential.

These approaches are complements, not competitors. The future of data management is one in which diverse data lives in the systems that are best suited to hold it long-term, and these systems make the data available with all of the security, reliability, and scalability that enterprise computing demands. Equally important, these systems must cooperate so seamlessly that analysts, data scientists, and developers don't need to know which data lives where.

Data Liquidity

However, data equality creates new problems. With a greater diversity of data feeding a greater variety of algorithms, analytics, and applications, the need to convert data from one shape to another skyrockets. The answer is data liquidity—the ability to get the data you want into the shape you need with minimal time, cost, and risk.

The easiest way to see the value of data liquidity is in data discovery, where analysts combine imperfect, irregularly shaped datasets into new arrangements to obtain previously unobtainable perspectives on the relationships among customers, products, and warranty claims, for example. Because these mashups are made without the laborious process of building predetermined models, they're easily pushed aside when a new question demands yet another new combination.

But data liquidity doesn't just mean more flexibility for people. Algorithms benefit, too. New in-memory database technologies represent data in both a row format for writing new records quickly, and a column format for fast analytical queries, done simultaneously. Data-lab environments with close ties to production servers enable algorithms with new models and tweaked weightings to go into production in hours instead of weeks, putting new ideas to work faster. And moving extract-transform-and-load (ETL) jobs from overburdened warehouse environments to Hadoop clusters makes it more cost-effective to quickly convert data into new shapes.

> Additional abilities to visually profile, transform, and explore data bring the cost of asking a new question below the cost of failing to do so.

Achieving these improvements in isolation is good, but attaining them as part of a coherent architecture is better. Some of the workloads will be well suited for public clouds, such as machine-learning algorithms that identify and classify people, places, and things mentioned in forum posts.

Others, such as algorithms for one-to-one marketing promotions based on customer purchase history, are better done behind the firewall. But augmenting that algorithm with external data requires a hybrid cloud architecture that combines public cloud capabilities with corporate data centers reconfigured as private clouds.

Data Security and Governance

The challenge of every chief data officer is to maximize value creation from a company's data capital while ensuring that its use complies with applicable policies, regulations, and laws. Security, and its constant partner, governance, are necessary checks on data liquidity.

A data-centric approach to security still incorporates traditional tactics such as hardening the perimeter. But it also recognizes the realities of data capital. Data doesn't sit still; it moves around. It gets used in different systems, by different people both inside and outside the enterprise.

As a result, securing data-capital computing means bringing new technologies such as Hadoop and Spark, as well as new cloud environments, under supervision. Established authorization, access, auditing, and encryption—for data both at rest and in motion through the network—must extend to the entire data-capital computing environment. Meanwhile, during a time of ever-stricter regulatory requirements, it's increasingly important for companies to establish strong data-governance policies. But for many companies, that effort remains on the "to-do" list: According to recent Rand Secure Archive research, 44 percent of organizations polled had no formal data-governance policies; of those, 22 percent had no plans to implement one. (For more on securely managing and governing big data, see the Oracle white paper "Securing the Big Data Life Cycle.")

Characteristics of Enterprise Architecture for Data Capital

Data Equality

Capturing and keeping data in its original shape and format requires diverse data management technologies working together seamlessly.

Data Liquidity

Get the data you want into the shape you need for the task at hand with minimal time, cost, and risk.

Data Security and Governance

Authorization, access, encryption, and auditing—for data both at rest and in motion—must extend to the full computing environment.

New Capabilities from Data-Capital Computing

Data-capital computing, with its hybrid-cloud architecture, proliferation of new data management, integration, and analytic services, and collaboration between open-source and commercial software, takes a lot of investment to get right. But large organizations are taking the plunge because of the new capabilities the approach delivers.

- **Data-driven tailoring.** Machine-learning algorithms that vary their output based on new input can tailor pricing, inventory management, and fraud detection. But figuring out what works under which conditions and for how long requires plenty of experimentation. Data-capital computing makes these experiments cost-effective to run and modify on the fly.

- **Internal data marketplaces.** In addition to canned analytics for known questions, data capital-intensive companies offer internal data markets that satisfy spot demand for new data products. Visual interfaces to Hadoop data reservoirs help managers and data scientists alike to browse for datasets in the same way that they might shop at any e-commerce site.

Additional abilities to visually profile, transform, and explore data bring the cost of asking a new question below the cost of failing to do so.

- **Data nudges.** Embedded analytics in applications encourage users toward desired outcomes. These include next-best actions for call-center reps, recommendations on e-commerce Web and mobile sites, and progress metrics in mobile apps that encourage people toward particular goals.

- **Data services.** Data can also be treated as a service. Some companies already create ways for applications to grab data on demand for the tasks they automate. These application programming interfaces, or APIs, cut the cost of using existing stocks of data capital in new ways, increasing the potential value of data.

Many companies collect data that they can share with others, after scrubbing proprietary and personally identifiable information. Some of that data may be helpful to their trading partners or suppliers. Other data may prove valuable to completely unrelated entities.

Looking Ahead

Experts predict that data's importance will only continue to grow.

The Internet of Things (IoT) will connect 6.4 billion devices worldwide in 2016 and exceed 20 billion intelligent, connected things by 2020, according to Gartner Inc. These devices are generating more and more data, Brynjolfsson notes. "Software is being embedded in more and more products, from autos to television to prescription medications," he says. "They're all generating data. That creates opportunities to analyze data, and that data analysis creates value."

But that's creating new challenges, as well. "Having data by itself is not sufficient," Brynjolfsson says. "If you get the data and it just sits there, that's not going to help you. Or if you use it in old-style decision-making, that's not going to help. You need to rethink your business processes about how you make decisions."

Historically, managers relied on "gut instinct" for decision-making simply because they lacked the data to do otherwise, says Brynjolfsson, whose co-authored 2012 *Harvard Business Review* article "Big Data: The Management Revolution" explores the issue in more depth. "Today, it's more scientific, and many managers are not accustomed to making decisions this way. It's a whole new culture."

In addition to sharpening the power of human judgment, data capital will extend the reach of algorithmic decision-making. As managers incorporate more data into their thinking, more kinds of decisions will reveal their inner logic. Choices once thought to be strictly the domain of human beings will become the province of machines. This already happens with marketing recommendations, delivery truck routing, and credit approvals. Areas that are still experimental today, such as algorithmic analysis of cancer-screening images, will become routine tomorrow.

How to Start—and How Oracle Can Help

Generating return from data capital is not simply a matter of adding new technology to the enterprise. It's a question of integrating that technology with the existing enterprise architecture to create sustainable competitive advantage. Oracle's comprehensive portfolio of big-data solutions, built on a hybrid cloud architecture, can help drive the competitive strategy and business value described in this paper.

One of the simplest ways to begin experimenting with new value from data is through cloud services. For example, Oracle Data Cloud's data-as-a-service mixes data on more than 200 million unique IDs from Oracle's BlueKai and Datalogix with a company's own customer data to create unique ad-targeting or remarketing campaigns. Oracle's IoT Cloud Service provides secure connections to any data-generating device, analytics for functions such as preventive maintenance and asset tracking, and integration into enterprise applications.

Other companies have begun exploring the potential value of their own data capital by using Oracle Big Data Discovery running on the Oracle Big Data Appliance with Cloudera Hadoop to create internal data marketplaces. That same Hadoop cluster can also play host to Oracle R Advanced Analytics for Hadoop, a version of the popular open-source statistical package modified to take advantage of enterprise-scale computing resources.

Companies that have proven the value of data discovery and mash-up environments want to run big-data workloads with the same service-level agreements, security, and disaster recovery as the rest of the enterprise environments. They look for enterprise-grade big-data management, such as connecting the Oracle Big Data Appliance to Oracle Exadata running Oracle Database 12c. Pre-built connectors and Oracle Big Data SQL make the two environments work together as one unified, high-performance system so that managers and data scientists alike can use the data they care about without having to know which system it's in.

Big-data integration technologies such as Oracle GoldenGate for Big Data make it easier and faster to expand the stock of data capital in the big-data management system. And tools such as Oracle Data Integrator for Big Data speed up and simplify the process of reshaping diverse data for a variety of endpoint algorithms, analytics, and applications.

Bottom line: If you demand to know the full value of your data capital before investing in it, you'll find some of that value fading away. So whether you're just beginning to bring these technologies into the business or looking to make them business as usual, don't hesitate to take the next step.

MIT TECHNOLOGY REVIEW CUSTOM

Produced in partnership with

ORACLE®

www.oracle.com/bigdata