CHAPTER 2

Leveraging an Expanding Universe of Data

EBay is a multibillion-dollar corporation with operations in more than 30 countries. Founded in 1995, the San Jose–based company is an early Internet pioneer that has managed to grow through changing markets and economic conditions to become the global e-commerce powerhouse it is today. Given such long-term resilience and success, it's easy to forget the considerable growing pains eBay suffered in the mid-2000s. That's when the company's core auction business began to plateau, putting eBay at a crossroads.

"Is eBay becoming a mature company that won't be able to keep up its heady growth?" asked a Wharton School report in February 2005, amid disappointing stock performance and a reduced financial outlook for that year. An October 2007 *Fortune* article, meanwhile, chronicled "eBay's transition to adulthood" as the company sought more growth through acquisitions and weathered a shift in how analysts valued the company—from web-centric statistics like new users and total auction listings to retail-industry metrics like overall sales growth and revenue generated per user.

These two citations are representative of many from the mid-2000s, and they fall within the tenure of the practitioner in your practitioner/academic coauthor team (Oliver), who served in senior analytics roles at eBay from November 2004 through October 2011. It was a time when analytics suddenly became a target for further growth and investment internally, as eBay realized the valuable role data and data products could play as a chief differentiator moving forward. eBay had begun collaborating on advertising with partners like Yahoo! and Google, for example, and those partners saw huge value in eBay's traffic and customer data. Analytics, in other words, became a top-line business driver almost overnight.

We'll share details later in this chapter on how some clutch moves by the eBay analytics team during this crucial period helped revitalize the company and laid groundwork for some core elements of the Sentient Enterprise approach. But first we need to recognize how these early efforts to prioritize analytics came in response to industry-wide changes that are still happening today.

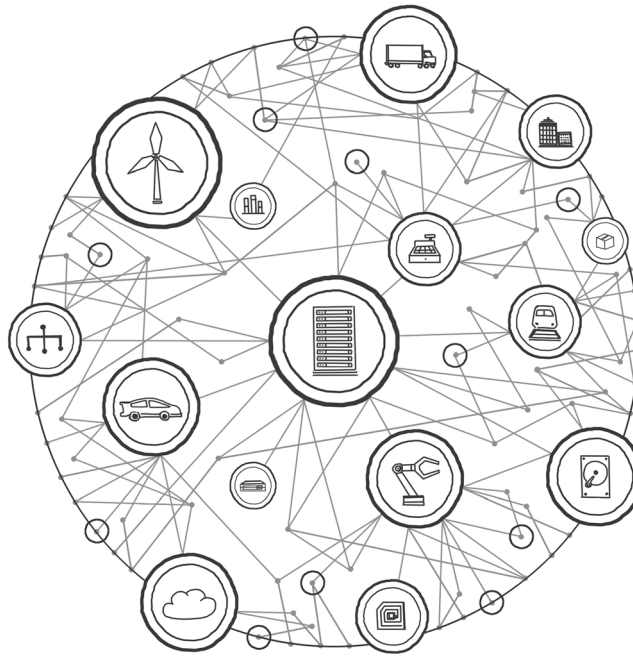
A UNIVERSE OF DATA: EXPANDING EXPONENTIALLY WITH NEW SOURCES

Embracing analytics as a market differentiator is now a rite of passage for any business reaching maturity or hoping to stay competitive in today's thoroughly data-driven economy. And analytics will only get more valuable as we find more—and more diverse—forms of data to fuel our inquiries. Indeed, the exponential growth in data volumes, complexity, and velocity makes the consequences and payoffs from how we leverage analytics all the more steep.

Research from the marketing intelligence firm IDC shows our digital universe is expanding 40 percent per year and is expected to hit 44 zetabytes (one zetabyte is a trillion gigabytes) by 2020. That's more than all the stars in the universe. A recent Cisco Visual Networking Index, meanwhile, projects more than 10 billion mobile-ready devices and connections by 2018, with average mobile connection speeds doubling by then to 2.5 megabytes per second. These are some of the dynamics behind the fact that 1.7 megabytes are being generated per minute for every person on Earth. But that doesn't mean all that data is being generated by people.

It turns out that online shopping behavior and other forms of human-generated data—like when someone types on a keyboard, takes a picture, hits the "record" button, or scans a bar code—are dwarfed by the so-called Internet of Things (IoT), made up of data from electronic sensors and other machine-generated sources. These "things" can be environmental sensors monitoring weather, traffic, or energy grid patterns; telemetry from machines, trucks, or store shelves to track manufacturing and distribution channels; and wearable devices that relay data about your health, location, and activity level.

Those examples are just the tip of the iceberg. With the IoT figuring prominently among its Top 10 Strategic Technology Trends, Gartner forecast some 26 billion IoT units installed by 2020 and IoT product and service suppliers generating more than $300 billion in revenue and $1.9 trillion in global economic value. IDC's research, meanwhile,



shows that sensor signals from embedded systems—a major IoT component—currently make up only 2 percent of the digital universe, but will reach 10 percent by the end of the decade.

The IoT continues to index our world on an unprecedented, granular level. As capacity grows, so do the chances to reap—or miss out on—immense value for your organization through analytics. The key to navigating our world's growing digital footprint lies in all the interactions and behaviors we see happening amid all that information.

In its simplest forms, analytics can achieve the straightforward goal of capturing transactions for a specific purpose. From there it can lead to: "Now that I've got those transactions, what else can I learn, and how else can I optimize?" With the IoT, you're suddenly dealing with a wider and more complex variety of data; and that means the possibilities grow exponentially as you figure out new ways to apply that data for analysis and learning.

Take the example of fleet logistics. The immediate value of a GPS sensor in a tractor trailer includes answers to questions like: "Where's my truck? Did my driver take too long a break or did he stop

somewhere that's not on the scheduled route?" That immediate interest is very transactional. But look subsequently at your fleet GPS data in aggregate over time, and you are led to some very interesting use cases for saving time and fuel through efficient route processing and other logistics.

Sensors in wind turbines, meanwhile, record efficiency ratios and other metrics for immediate transactional needs like repair or warranty issues. But keeping track from an analytics perspective can also help you leverage that information as input for future decisions around design criteria or placement of new turbines.

The growing list of use cases includes everyday examples like the so-called black-box event data recorders in most new cars, with metrics around driving habits and vehicle performance that can serve transactional needs like fuel efficiency and crash investigations, as well as predictive needs like engine and safety design modifications in new models. And passenger cars are increasingly able to automatically capture and transmit diagnostics for future purposes like scheduling maintenance and repairs, as well as real-time needs.

For instance, one of us (Mohan) had trouble starting a Tesla in the driveway on a very cold February morning. Thanks to remote diagnostics systems that deliver real-time information, one call to the service center allowed a service representative to immediately log into the car and benchmark its power consumption against that of other Teslas within a 50-mile radius. Pinpointing the problem as cold starts, the technician advised preheating the battery for a few minutes, something easily done via app from inside the house. That's the future of service in an IoT world: zero-touch and real-time repairs!

Cars may one day be able to order their own replacement parts, or even connect with a consumer's 3D printer to create the part at home in the garage. Especially when we combine multiple existing technologies—black-box sensors and 3D printers in this case—get ready to see entire business models redefined in radical ways.

Of course, black boxes have been in airplanes a lot longer than they've been in cars; the devices (which are actually colored orange) have been mandatory in U.S. commercial jets since 1967. Today's sensor revolution, however, allows modern black boxes to assess airplane and crew behavior with unprecedented clarity. And now The Boeing

Company is leading an effort to increase connectivity among multiple black boxes! Boeing has filed patents for a system that would link the data from many black boxes, helping give aviation experts a granular and fleetwide understanding of how planes are operating and where problems may lie.

Not surprisingly, Siemens is another company that's pioneering connectivity, data integration, and analysis on an unprecedented scale. Founded in 1847, the global engineering powerhouse employs more than 340,000 people worldwide, with annual revenue upwards of $86 billion. The company recently implemented a common technology base that processes more than 17 terabytes of data per month for remote monitoring and maintenance of more than 300,000 connected devices—everything from traffic light controls and locomotives to computed tomography (CT) scanners and the sensor-laden wind turbines we mentioned a few paragraphs back.

Gerhard Kress is vice president of data services at Siemens Mobility, the division that focuses on railway vehicles and control systems. "The sensor data from just one fleet of trains in Europe can fill about 100 billion lines of a table," he told us. "The right analytics on all this data can let us know—a week or two in advance—that a component on a train is going to fail, and we can take steps to stop or minimize the problem before it happens. That's something we couldn't have done just a few years ago." (We'll have a closer examination of Siemens's groundbreaking success at predictive maintenance in Chapter 7.)

Whether the goal is improving reliability, boosting efficiency, or saving lives, it's pretty powerful and exciting to see how a combination of human-generated data and the IoT can help us detect learned behaviors and patterns from various data sources to guide proactive interventions and future decisions in very specific and accurate ways.

GAME-CHANGING CAPABILITIES

We've stressed how data is transforming countless industries for competitive advantage. Perhaps nowhere is this analytic quest for competitive advantage taken more literally than in our national pastime: baseball.

Perhaps you're familiar with *Moneyball*, the data-driven Cinderella story about how the 2002 Oakland Athletics used player statistics and algorithms to build a winning baseball team. From an analytics perspective, the closest your coauthors ever got to "inside baseball" on the subject came by way of a 2016 technology summit at a seaside conference center in Half Moon Bay, California. One session in particular, featuring Bill Schlough, senior vice president and chief information officer for the San Francisco Giants, made clear how far the sport has come since the early days of *Moneyball*.

For instance, when *Moneyball* innovator Billy Beane first put analytics in service for Oakland, he was focused mostly on scouting talent through the use of relatively small sets of historical data. Most of it was high-level summary data about players' past performance. Today, advanced radar can help make critical decisions in the middle of the game.

Sensors track not just the speed of a ball, but also the arc and spin to help determine if a pitcher is getting fatigued. "You can use this data to detect when the rotation on the ball is starting to slow—it's not breaking as much as it was before on previous pitches," said Bill Schlough. "That's a sign that your pitcher may be about to give up a run; and that's the kind of information the coach can use to pull him from the mound before that happens."

At the point we caught up with Bill in October 2016, the Giants were enjoying a 490-home-game sellout streak at the team's 41,000-seat stadium.

While Bill had to be careful not to discuss competitive advantage, our time together with him revealed enough for us to understand how one of the most successful franchises in baseball also happens to be heavily driven by analytics.

"Willie Mays was once asked about the pitcher he feared most, and his answer was, 'Show me a pitcher I've never seen.' That's why, when we host another team, our sensors track how their pitchers throw," Bill said. "So, what's to stop you from programming those dynamics into a pitching machine?"

The Giants also embrace technology to optimize engagement with entire demographic groups among its fan base. Take the case of

millennials. "No millennial is going to sit still for a three-hour game with two hours and 40 minutes of dead time and 20 minutes of action," Bill said. "That's why we have full Wi-Fi connectivity and TVs in the lounge areas, so it's like a huge sports bar where they can follow multiple games."

WELL-INTENTIONED ANARCHY

For all its promise across diverse industries and applications, today's massive flood of information can reap top value only if you embrace architectures that support the kind of workplace collaboration we talked about in this book's Introduction. Otherwise, we're stuck in a reactive and very confusing situation. Consider a story that perhaps you can relate to:

> A couple of weeks before the Q4 earnings call, the CFO of a telecommunications company receives sales and revenue numbers for the quarter and notices a five percent dip in revenue. She calls her executive team together and assigns them the task of trying to figure out why. Her team is divided on whether the number is even accurate. Those analysts who agree with it suggest that this dip is just fallout from the problems that the company started having with service issues back in January when several snowstorms hit major East Coast and Midwest urban areas hard.
>
> The company was never able to fully recover from angry customers canceling their contracts, and analysts are not really sure why the company kept losing them, even in the sunny summer months. These problems, compounded by a decline in discretionary consumer spending and a competitor's new contract with the iPhone, are only now just becoming apparent. The analysts sigh and say this might just have to be the status quo everyone needs to accept.

Unfortunately, this scenario is all too common. It depicts a company that's reactive instead of proactive, and is far too sluggish in identifying small problems before they snowball into catastrophic ones.

It turns out that more than half (57 percent) of the 362 global executives in a recent survey by the Economist Intelligence Unit said that important business data is not captured or shared at their companies. Furthermore, 42 percent said that their data is cumbersome and not user-friendly.

True agility requires us to break down silos so that a variety of professionals from across the organization can work together around data to unlock new insights. At the same time, we need to be careful in how we break down those silos and what kind of governance we put in their place, or else we'll suffer through the analytics pain points mentioned in Chapter 1. Yes, we need to make data accessible. But we can't be very agile if doing so leaves our organization swamped in a mess of incoherent data—a good deal of it duplicated or wrong because we don't have the systems or governance to manage it well.

DATA MARTS AND THEIR DISCONTENTS

Remember our scenario in Chapter 1, where our midlevel data scientist didn't feel like waiting a year and a half for an official IT solution? He got resourceful and built his own data mart, a business-unit-specific silo that copies the data he needs so he can get his project done more quickly. Unfortunately, his rock-star solution is throwing rocks into the system. Here's why:

Data marts create data drift, which happens when data loading and maintenance aren't coordinated. This results in mismatched structures and redundancy that lead to inconsistent answers. Let's assume, for instance, that you have the CMO controlling one data mart and

the CFO controlling another, and they're trying to pull the same metric—let's say the metric is the number of active customers this month. They'll likely come up with very different numbers because of misaligned databases, hardware, software platforms, and reporting tools that make it difficult or impossible to standardize and integrate information.

In our own experience, we've seen how data marts can lead to unnecessary duplication by a factor of 20 times or more, with data drift of 60 percent or more. The costs here are not abstract, and they go beyond the capacity and memory for all that unnecessary duplication. Imagine that the skewed data your CFO and CMO are working with involves promotional materials targeted to VIP customers. How much time, effort, and money are going toward sending marketing collateral and special offers to the wrong customers?

Given these costs, the redundant data operations, and hours of discussion about how to recalibrate misaligned metrics, we've gotten used to saying—only half-jokingly—that "data marts simply can't be cheap enough to justify their existence." Nothing about them is agile; it's the Wild West, and these data silos are Public Enemy Number One.

This is all a troubling picture not just because of the bad data and lost market opportunities that result, but also because of the effect it has on workforce culture and performance. Think about it: We hire people to be intelligent and proactive and to get things done. We pay recruiters top dollar to get folks who think outside the box and innovate through challenges. These are the qualities that drive success. Unfortunately, that gut feeling of success and agility the go-getters have when building a data mart is just a mirage; the reality is that the organization has actually become less agile because of data anarchy and ballooning IT costs to handle all the redundancies.

The road to data anarchy may be paved with good intentions, but it still leads us to the point where the majority of people's time can be spent on the fallout from this setup. Think of the daily or weekly fire drills that take place when numbers don't match up and the circular arguments begin over which data is right and which data is wrong. Users and technology executives alike can get stuck spinning their wheels if the organization doesn't have systems and policies that are inclusive and effective around data.

Business users sequestered from the analytic tools they need tend to either give up or go it alone through stovepipes and silos. It's no wonder that user frustration can arise under these circumstances. In *Drive*, his best-selling book on workplace motivation, author Daniel Pink explains how scientists have developed a "new operating system" for business success that revolves around three elements: autonomy, or the urge to direct our own lives; mastery, the desire to get better and better at something; and purpose, the yearning to do something that matters. It's not hard to look at these insights and see how most current IT policies and methodologies come up woefully short.

A SOLUTION? "LINKEDIN FOR ANALYTICS"

We talked in Chapter 1 about the rapidly changing expectations around data, with users from all walks of life now demanding real-time interaction. And they expect low latency; a Google study, in fact, showed it takes only 250 milliseconds for a user to get tired of waiting for a site to load. That's literally less than the blink of an eye! These factors are driving a new self-service ethic where everyone expects to be able to access data quickly and in ways that suit his or her own needs. Your own employees have these expectations as well.

Standing up the technological systems in the workplace to meet these internal expectations is tough, but business users don't care! When you tell members of your finance or marketing department it may take up to 18 months to complete an intensive data research project, don't be surprised when they set up their own data mart. They don't care that the supposed agility they perceive is actually polluting the larger analytics environment with data anarchy that leads to inconsistent or just plain wrong answers.

To keep business users aligned and engaged, we must tailor the analytics to them. For this, we borrow lessons from social media, gaming, and other areas where people naturally—even compulsively—want to take part.

It's important to preview a crowdsourcing approach, "LinkedIn for Analytics," we'll learn more about in Chapter 5. It's not related to the LinkedIn company, just inspired by that social media platform. The idea is to bring to analytics that same culture of engagement you see

on LinkedIn and many other social platforms or gaming environments. And it solves some key concerns about scalability in the process.

If you have a fairly small operation, you may be able to survive with a traditional approach like having your centralized team of analysts assign metadata so the rest of the company knows what information is important and where to find it. But in large organizations dealing with big data today, that traditional approach can quickly break down. Humans don't scale the way data does, and a corps of a hundred or even a thousand analysts still won't be able to keep up with the job of documenting the huge volumes of information and lightning-fast data streams coming at them.

That's why we need to turn to the wisdom of the crowd; specifically, the hundreds or thousands of people within your organization who work with data. At its core, LinkedIn for Analytics is essentially analytics on your analytics community. It starts with algorithmic models to examine what your community of data scientists and other analysts is doing with data, but we're not simply interested in their specific queries or dashboard activity. We want to provide a forum to capture and analyze commentary and discussion between these folks, complete with social media conventions to let people "like" a certain analytic approach, "follow" a particular analyst, or monitor visualizations and data sets that are popular and trending.

Suddenly, these patterns you see in the kinds of ideas, projects, and people that get followed, shared, and liked help you answer questions such as: "Who are the influencers?" "What projects and ideas are gathering the most energy?" "What does this tell us about the most important projects and their potential success outside the organization?" These insights open up infinite possibilities for innovation within your company.

We're not the first to crowdsource technical innovation by leveraging inherent human tendencies. The U.S. Department of Defense sponsors an online "shredder challenge" contest, with a $50,000 prize for whoever can best reconstruct documents that have been mercilessly shredded, with the shards then posted online. In another case, medical researchers used a gaming approach to help map a crucial AIDS enzyme (think Tetris for a good cause).

The end result is a business user community aligned and engaged in the task of tailoring and safely experimenting with data around the business problems they care about most. And just like the rest of the social media world, the best solutions go viral.

GETTING BACK TO EBAY: FULFILLING THE ANALYTICS MANDATE

By now, you probably see how any practical and scalable implementation of next-generation analytics in the enterprise demands we strike a careful balance between agility, governance, and a culture of inclusion around data. In the next handful of chapters that deal with each of the five stages of the Sentient Enterprise capability maturity model, we'll show in detail how to put these principles into practice at your company today.

Back the mid-2000s, however, these principles were just coming into focus for the eBay analytics team as we tried to make good on our mandate to stand up architectures commensurate with the newly critical role analytics needed to play in our company's survival. Some of these principles, in fact, became clear only in hindsight.

While our team was highly motivated and the company brass was on board with the effort, we still found ourselves at square one in many ways when it came to working with eBay's business users.

Quite frankly, the subjective experience was that our processes and methodologies had given us a bad reputation, and we weren't seen internally as good partners. In many ways, we had some of the most capable systems on Earth; but business users were running away from us for all the reasons we've laid out in this book so far about the requirements and ticket-driven status quo.

We realized we had to deliver solutions that not only helped the company, but also fit into the ways people naturally wanted to work so that IT collaboration stopped seeming like a chore for them. We chose a few key projects to turn things around.

One effort involved setting up collaborative platforms that became known first as data labs and, later, as the DataHub. A precursor to the Sentient Enterprise's LinkedIn for Analytics approach today, the

DataHub we created in the mid-2000s allowed eBay analysts to post interesting discoveries and correlations, with other eBay colleagues able to query the DataHub for tagged entries based on their areas of interest. People could add comments, follow a discussion, or link to certain groups focused on a specific tool or initiative. We designed the DataHub to allow safe experimentation; users could not just read analyst reports, but also manipulate charts and graphs as needed without messing up the data itself.

On another front, the eBay analytics team stepped up to help their vice president of Internet marketing sharpen the company's online advertising. Specifically, we needed to build analytics and instrument our website to capture and trace customer activity and outcomes on a very minute level—clickstream and other granular activity involving the behavioral data that we'll discuss more fully in Chapter 4. In the process, we reaped better insights than you could ever get from conducting customer surveys and similar customer relationship management (CRM) research.

As background, most customers never complain or respond to surveys; they just walk away and spend their money elsewhere if they're not happy. The analytics we built for the marketing team helped look beyond what customers say (or don't say) and examine their actual behavior with systematic precision. This can clarify friction points in the customer experience, and, in our case, it helped us understand overall customer behavior and preferences so we could use that knowledge to sharpen our targeted online marketing.

A final eBay example to share involves an analytics solution to optimize efficiencies in our data infrastructure around the world. We had built our reputation around capacity and reliability; now we needed analytics to preserve all that while boosting efficiency behind the scenes. To do this, we looked beyond immediate operations and metrics and charted performance over a three-month period.

This data allowed us to optimize server usage (we made changes in server deployment, load balance, and data traffic) to make operations more efficient to the point where 12,000 servers were able to do the work of what 15,000 servers had done before. Keep in mind this was back in 2008, right when the economic recession was hitting.

Analytics helped position eBay to survive that downturn by fulfilling capacity with a fraction of the infrastructure we had once needed.

We hope these examples show how taking an agile and collaborative approach to designing and building analytics architectures will help your company survive and compete in constantly changing and highly competitive markets. We were particularly pleased to see that eBay itself now references the Sentient Enterprise in an "Extreme Analytics @ eBay" presentation the company has been delivering at major events and conferences.

In the chapters to come, we'll use real-world examples like this wherever possible to explain what happens in the Sentient Enterprise's five stages, and how these innovations are already making a difference for real companies. The Sentient Enterprise is all about harvesting insights from experimentation, pilot projects, and lessons learned so we can understand and replicate agile success stories across industries—and at scale.

Agile and at scale. Bringing those two concepts together is, indeed, our key priority in presenting this book. You'll find plenty of books about agile ventures and entrepreneurial projects; there are also plenty of books on the shelf dealing with large-scale organizational management issues. In putting forth *The Sentient Enterprise*, we are bringing the two worlds together in ways you can model and replicate for success in your own company. The next five chapters (3 to 7) serve as your road map through the five stages of this capability maturity model. Our first stop is the Agile Data Platform.