This article was published by CNBC on November 6, 2014

Warning: Big Data Could Mean Big Risk

CEOs may understandably feel whipsawed these days between the need to keep up in the big data arms race and the risks of protecting that data from a breach.

Data-security breaches show no signs of abating in magnitude or frequency. In what may turn out to be the biggest breach of a retailer's computers ever, JPMorgan Chase confirmed at the beginning of this month that hackers had breached as many as 76 million customer accounts. That means that five of the six biggest data breaches of all time happened within the past 18 months. EBay, for example, recently announced that the personal information of 145 million customers had been breached. Last year alone, Adobe, Target, and Evernote suffered data breaches affecting 152 million, 110 million, and 50 million customers, respectively.

Meanwhile, the big data revolution has spread from the technology sector, financial services, and health care to virtually every industry, redefining business models and the terms of competition. And, as Tom Davenport, Babson College professor of Management and Information Technology, has argued, the application of big data now encompasses not only powerful data collection and analytics to support operations but also the embedding of "smartness" in products and services. Companies that remain on the sidelines or move too slowly are likely to see themselves far outdistanced by competitors.

But as every business leader knows and deals with as a daily part of the job, doing business requires taking risks. And the bigger the data, the bigger the risk: the more ways it can be compromised, the more valuable it is, and the more desirable it is as a target. Business leaders making decisions about their companies' big data proposals can't avoid those risks, but they can mitigate them by obeying these principles:

* **Be aware that somebody somewhere understands your data better than you do.** Just because it's your company's data, about your customers, you're not necessarily the expert on what inferences can be drawn from it and what crimes can be cooked up from it. Today's most advanced cybercriminals are looking not just for commodities like credit-card information, but for high-value inferences or other intellectual property that they can use for more sophisticated purposes.

For example, if you have ads on your web pages, chances are good that you engage a third party to manage those ads. This third party also has other clients, and so is in a good position to

aggregate data about your customers across a number of companies. The third party now owns better information about your customers' online behavior than you do. Couple this with the latest trend, called "device fingerprinting," and aggregators can develop profiles even for visitors who take steps to guard their privacy (including plugins that block cookies, regularly deleting cookies, or even routing through anonymizer sites first). And they can sell that information. But as disturbing as this may sound, selling customer behavioral profiles culled from various sources is really just another competitive intelligence (CI) challenge. And viewed this way, you probably already have staff or a service that manages CI threats through a combination of privacy protection, detection, and risk reduction mechanisms — so this new threat variation can be added to their scope.

*** Decide how you intend to use information to shape the business** – its operations, business model, or products and services – and only then collect data. Simply stockpiling data first, as many companies have done, and then combing it for insights adds up to an infinite risk profile. Consider: if you already know the key metrics that measure customer satisfaction, operational efficiency, or opportunities for improvement, then you don't need to stockpile all data to act on that knowledge. And if you lack that knowledge, stockpiling all data hoping to find it will result in, well, a pile of data attractive to hackers. You have a much better chance of determining with some accuracy the risk/reward ratio when you have carefully circumscribed your business objective before seeking out the data.

A major automaker, wishing to strengthen customer relationships, asked customers a single question: "How easy are we to do business with?" Responses were ranged along a scale of 1-6 and the company concentrated on those in the middle, realizing that the very dissatisfied were lost opportunities and the highly satisfied already well taken care of. This simple approach, based on knowing what they were looking for, enabled the company to jettison irrelevant data and the risk that went with it.

*** Bring to big data proposals the same business reflexes you have applied to highly complex technical challenges.** Sweeping big-data proposals often leads a familiar dilemma: whether to adopt a gold-plated solution or a solution that "merely" suffices. It's not a technical decision but a business decision of the kind that you have probably faced many times before.

For example, when PCs first appeared, the high priesthood of mainframe computing proclaimed them toys. Instead of being intimidated by the experts, savvy business leaders decided in favor of desktops, which then became ubiquitous in industry. A similar drama played out with spreadsheets. Business leaders chose them over far more expensive decision-support tools because they provided answers that were good enough, though not exhaustive or perfect. Just as you don't have to be an IT expert to make wise decisions about technology, you don't have to be a data scientist to make wise decisions about big data.

*** Guard your back doors, too.** Often the back door of a home is not as secure as the front door, and burglars know it. The same goes for the company data system — and it has hundreds of

backdoors. The front door — the "live" system used in daily operations — is usually well secured by multiple layers of defense. The back doors — the many places inside the company where that same data is stored or manipulated — are often less secure. And there are many such places: a disaster recovery system that maintains an exact copy of the live system, numerous test environments where the next iterations of the system and subsystems are being tested, and even more numerous development environments (sometimes on individual laptops or offshore) where developers pull down data from the live system.

"De-identification" provides one leading-edge solution for these environments. It safeguards data by jumbling the identity of individuals and other sensitive information about them. But it's hard to do right, and requires some judgment — after all, the very data correlations you break to render the data safe may be the same subtle correlations your software wants to leverage for market value. This is a good time to consider widely encrypting your data. The complexities are fairly well known, and the protection you gain is robust.

Is big data rocket science? In the details, yes. But you don't have to be a rocket scientist to put it to work safely and successfully. You just have to practice smart analytics, not big data for big data's sake. And approach the task incrementally, step by step, guided by your own experience in whittling away at large risks until they are acceptably small.

Commentary by Glenn Kapetansky, chief security officer for Trexin Consulting, a management- and technology-consulting firm specializing in the application of advanced technologies that drive business value. Follow the company on Twitter @trexin.
