

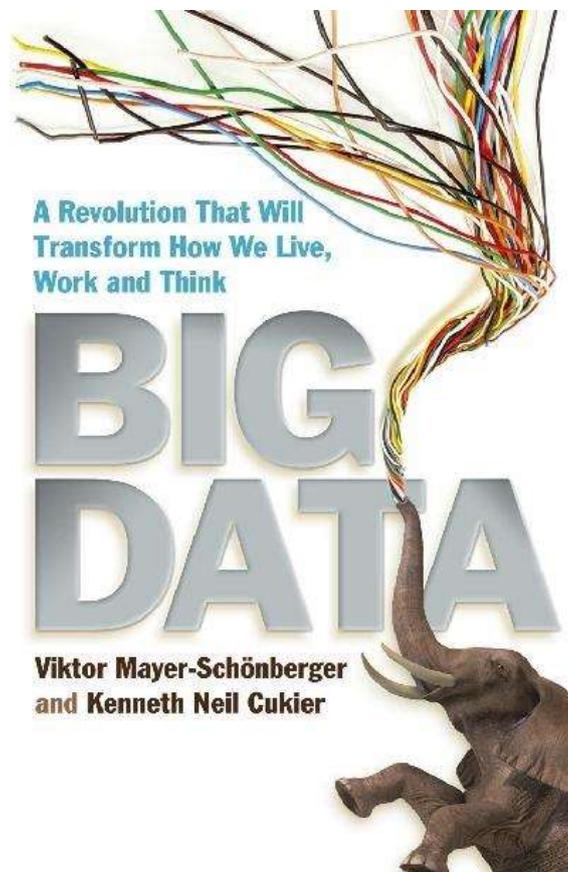
A summary of the book

# Big Data

A revolution that will transform how we live, work and think

By Victor Mayer-Schönberger & Kenneth Cukier

Summary by Kim Hartman



This is a summary of what I think is the most important and insightful parts of the book. I can't speak for anyone else and I strongly recommend you to read the book in order to fully grasp the concepts written here. My notes should only be seen as an addition that can be used to refresh your memory after you've read the book. Use the words in this summary as anchors to remember the vitals parts of the book.

More book summaries at [www.kimhartman.se](http://www.kimhartman.se)

Contact me at [Kim.hartman@hyperisland.se](mailto:Kim.hartman@hyperisland.se)

## Contents

Description from amazon.....	2
Chapter 1: Now.....	3
Chapter 2: More .....	4
Chapter 3: Messy .....	6
Chapter 4: Correlation .....	7
Chapter 5: Datafication.....	9
Chapter 6: Value .....	12
Chapter 7: Implications.....	15
Chapter 8-9: Risks/Control.....	18
Chapter 10: Next.....	19
Previous book summaries .....	21

More book summaries at [www.kimhartman.se](http://www.kimhartman.se)

Contact me at [Kim.hartman@hyperisland.se](mailto:Kim.hartman@hyperisland.se)

## **Description from amazon**

A New York Times bestseller. Long listed for the Financial Times/Goldman Sachs Business Book of the Year Award. Since Aristotle, we have fought to understand the causes behind everything. But this ideology is fading. In the age of big data, we can crunch an incomprehensible amount of information, providing us with invaluable insights about the what rather than the why. We're just starting to reap the benefits: tracking vital signs to foresee deadly infections, predicting building fires, anticipating the best moment to buy a plane ticket, seeing inflation in real time and monitoring social media in order to identify trends. But there is a dark side to big data. Will it be machines, rather than people, that make the decisions? How do you regulate an algorithm? What will happen to privacy? Will individuals be punished for acts they have yet to commit? In this groundbreaking and fascinating book, two of the world's most-respected data experts reveal the reality of a big data world and outline clear and actionable steps that will equip the reader with the tools needed for this next phase of human evolution.

## Chapter 1: Now

**Big data:** the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value. Big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more.

**The raw material of business:** Rather, data became a raw material of business, a vital economic input, used to create a new form of economic value. In fact, with the right mindset, data can be cleverly reused to become a fountain of innovation and new services. The data can reveal secrets to those with the humility, the willingness, and the tools to listen.

**A change of state:** Half a century after computers entered mainstream society, the data has begun to accumulate to the point where something new and special is taking place. Not only is the world awash with more information than ever before, but that information is growing faster. The change of scale has led to a change of state. The quantitative change has led to a qualitative one.

**The revolution is in the data:** The real revolution is not in the machines that calculate data but in data itself and how we use it. The amount of stored information grows four times faster than the world economy, while the processing power of computers grows nine times faster. A movie is fundamentally different from a frozen photograph. It's the same with big data: by changing the amount, we change the essence.

**Predictions:** At its core, big data is about predictions. Though it is described as part of the branch of computer science called artificial intelligence, and more specifically, an area called machine learning, this characterization is misleading. Big data is not about trying to "teach" a computer to "think" like humans. Instead, it's about applying math to huge quantities of data in order to infer probabilities. Big data is about what, not why. We don't always need to know the cause of a phenomenon; rather, we can let data speak for itself.

**Macro instead of micro:** As scale increases, the number of inaccuracies increases as well. With big data, we'll often be satisfied with a sense of general direction rather than knowing a phenomenon down to the inch, the penny, the atom. We don't give up on exactitude entirely; we only give up our devotion to it. What we lose in accuracy at the micro level we gain in insight at the macro level.

**The impact of Big Data:** Big data changes the nature of business, markets, and society. In the twentieth century, value shifted from physical infrastructure like land and factories to intangibles such as brands and intellectual property. That now is expanding to data, which is becoming a significant corporate asset, a vital economic input, and the foundation of new business models.

## Chapter 2: More

**Big data:** is all about seeing and understanding the relations within and among pieces of information.

**Data that speaks:** The digital age may have made it easier and faster to process data, to calculate millions of numbers in a heartbeat. But when we talk about data that speaks, we mean something more—and different. As noted in Chapter One, big data is about three major shifts of mindset that are interlinked and hence reinforce one another.

1. The first is the ability to analyze vast amounts of data about a topic rather than be forced to settle for smaller sets.
2. The second is a willingness to embrace data's real-world messiness rather than privilege exactitude.
3. The third is a growing respect for correlations rather than a continuing quest for elusive causality.

**Randomness:** Statisticians have shown that sampling precision improves most dramatically with randomness, not with increased sample size. Random sampling has been a huge success and is the backbone of modern measurement at scale. But it is only a shortcut, a second-best alternative to collecting and analyzing the full dataset. It comes with a number of inherent weaknesses. Most troublingly, random sampling doesn't scale easily to include subcategories, as breaking the results down into smaller and smaller subgroups increases the possibility of erroneous predictions.

**Using ALL data:** After a certain point early on, as the numbers get bigger and bigger, the marginal amount of new information we learn from each observation is less and less. Using ALL the data makes it possible to spot connections.

**Sampling:** Sampling quickly stops being useful when you want to drill deeper, to take a closer look at some intriguing subcategory in the data. What works at the macro level falls apart in the micro. Sampling is like an analog photographic print. It looks good from a distance, but as you stare closer, zooming in on a particular detail, it gets blurry.

**Using entire dataset or as much data as possible:** But the absolute number of data points alone, the size of the dataset, is not what makes these examples of big data. What classifies them as big data is that instead of using the shortcut of a random sample, both Flu Trends and Steve Job's doctors used as much of the entire dataset as feasible. As when converting a digital image or song into a smaller file, information is lost when sampling. Having the full (or close to the full) dataset provides a lot more freedom to explore, to look at the data from different angles or to look closer at certain aspects of it. Big data relies on all the information, or at least as much as possible, it allows us to look at details or explore new analyses without

More book summaries at [www.kimhartman.se](http://www.kimhartman.se)

Contact me at [Kim.hartman@hyperisland.se](mailto:Kim.hartman@hyperisland.se)

the risk of blurriness. An investigation using big data is almost like a fishing expedition: it is unclear at the outset not only whether one will catch anything but what one may catch.

## Chapter 3: Messy

**Imprecision and messiness:** In many new situations that are cropping up today, allowing for imprecision—for messiness—may be a positive feature, not a shortcoming. It is a tradeoff. In return for relaxing the standards of allowable errors, one can get ahold of much more data. Treating data as something imperfect and imprecise lets us make superior forecasts, and thus understand our world better.

**Tags:** The imprecision inherent in tagging is about accepting the natural messiness of the world

**Example of measuring a vineyard with several data points:** We need to measure the temperature in a vineyard. If we have only one temperature sensor for the whole plot of land, we must make sure it's accurate and working at all times: no messiness allowed. In contrast, if we have a sensor for every one of the hundreds of vines, we can use cheaper, less sophisticated sensors (as long as they do not introduce a systematic bias). Chances are that at some points a few sensors may report incorrect data, creating a less exact, or "messier," dataset than the one from a single precise sensor. Any particular reading may be incorrect, but the aggregate of many readings will provide a more comprehensive picture. Because this dataset consists of more data points, it offers far greater value that likely offsets its messiness.

**The big picture:** We get a more complete sense of reality—the equivalent of an impressionist painting, wherein each stroke is messy when examined up close, but by stepping back one can see a majestic picture. Big data, with its emphasis on comprehensive datasets and messiness, helps us get closer to reality than did our dependence on small data and accuracy. Our comprehension of the world may have been incomplete and occasionally wrong when we were limited in what we could analyze, but there was a comfortable certainty about it, a reassuring stability. Besides, because we were stunted in the data that we could collect and examine, we didn't face the same compulsion to get everything, to see everything from every possible angle.

## Chapter 4: Correlation

**What and why:** Knowing why might be pleasant, but it's unimportant for stimulating sales. Knowing what, however, drives clicks. This insight has the power to reshape many industries.

**Correlations:** At its core, a correlation quantifies the statistical relationship between two data values. A strong correlation means that when one of the data values changes, the other is highly likely to change as well. A weak correlation means that when one data value changes little happens to the other. Correlations cannot foretell the future; they can only predict it with certain likelihood. But that ability is extremely valuable. Predictions based on correlations lie at the heart of big data.

**Predictive analytics:** Is starting to be widely used in business to foresee events before they happen. The term may refer to an algorithm that can spot a hit song, which is commonly used in the music industry to give recording labels a better idea of where to place their bets. Predictive analytics may not explain the cause of a problem; it only indicates that a problem exists. It will alert you that an engine is overheating, but it may not tell you whether the overheating is due to a frayed fan belt or a poorly screwed cap. The correlations show what, not why, but as we have seen, knowing what is often good enough.

**Nonlinear relationships:** Before big data, partly because of inadequate computing power, most correlational analysis using large data sets was limited to looking for linear relationships. In reality, of course, many relationships are far more complex. With more sophisticated analyses, we can identify non-linear relationships among data.

**Causality:** When we say that humans see the world through causalities, we're referring to two fundamental ways humans explain and understand the world: through quick, illusory causality; and via slow, methodical causal experiments. Big data will transform the roles of both. Also, we are biased to assume causes even where none exist. It is a matter of how human cognition works. When we see two events happen one after the other, our minds have a great urge to see them in causal terms. The fast-thinking side of our brain is hard-wired to jump quickly to whatever causal conclusions it can come up with.

**Correlations and causality:** Like correlations, causality can rarely if ever be proven, only shown with a high degree of probability. But unlike correlations, experiments to infer causal connections are often not practical or raise challenging ethical questions. Correlations are not only valuable in their own right, they also point the way for causal investigations. By telling us which two things are potentially connected, they allow us to investigate further whether a causal relationship is present, and if so, why. Through correlations we can catch a glimpse of the important variables that we then use in experiments to investigate causality. Correlations exist; we can show them mathematically. We can't easily do the same for causal links. So we would do well to hold off from trying to explain the reason behind the correlations: the why

More book summaries at [www.kimhartman.se](http://www.kimhartman.se)

Contact me at [Kim.hartman@hyperisland.se](mailto:Kim.hartman@hyperisland.se)

instead of the what. Non-causal methods based on hard data are superior to most intuited causal connections, the result of fast thinking.

**Data instead of hypotheses:** Big data transforms how we understand and explore the world. In the age of small data, we were driven by hypotheses about how the world worked, which we then attempted to validate by collecting and analyzing data. In the future, our understanding will be driven more by the abundance of data rather than by hypotheses. The traditional process of scientific discovery—of a hypothesis that is tested against reality using a model of underlying causalities—is on its way out, Anderson argued, replaced by statistical analysis of pure correlations that is devoid of theory. Because big-data analysis is based on theories, we can't escape them. They shape both our methods and our results. It begins with how we select the data.

## Chapter 5: Datafication

**Example of big data with cars:** Few would think that the way a person sits constitutes information, but it can. When a person is seated, the contours of the body, posture, and distribution of weight can all be quantified and tabulated. Koshimizu and his team of engineers convert backsides into data by measuring the pressure at 360 different points from sensors in a car seat and indexing each point on a scale from zero to 256. The result is a digital code that is unique for each individual. In a trial, the system was able to distinguish among a handful of people with 98 percent accuracy. The research is not asinine. The technology is being developed as an anti-theft system in cars. A vehicle equipped with it would recognize when someone other than an approved driver was at the wheel and demand a password to continue driving or perhaps cut the engine. Transforming sitting positions into data creates a viable service and a potentially lucrative business. And its usefulness may go far beyond deterring auto theft. For instance, the aggregated data might reveal clues about a relationship between drivers' posture and road safety, such as telltale shifts in position prior to accidents. The system might also be able to sense when a driver slumps slightly from fatigue and send an alert or automatically apply the brakes. Professor Koshimizu took something that had never been treated as data—or even imagined to have an informational quality—and transformed it into a numerically quantified format.

**Datafication:** There is no good term yet for the sorts of transformations produced by Commodore Maury and Professor Koshimizu. So let's call them Datafication. To datafy a phenomenon is to put it in a quantified format so it can be tabulated and analyzed. Again, this is very different from digitization, the process of converting analog information into the zeros and ones of binary code so computers can handle it. Measuring reality and recording data thrived because of a combination of the tools and a receptive mindset. That combination is the rich soil from which modern Datafication has grown.

**Digitization:** The IT revolution is evident all around us, but the emphasis has mostly been on the **T**, the technology. It is time to recast our gaze to focus on the **I**, the information. In short, digitization turbocharges Datafication. But it is not a substitute. The act of digitization—turning analog information into computer-readable format—by itself does not datafy. Information has stored value that can only be released once it is datafied.

**Google's Ngram Viewer:** <http://books.google.com/ngrams>

**Geo-location:** The geo-location of nature, objects, and people of course constitutes information. The mountain is there; the person is here. But to be most useful, that information needs to be turned into data. To datafy location requires a few prerequisites. We need a method to measure every square inch of area on Earth. We need a standardized way to note the measurements. We need an instrument to monitor and record the data. Quantification, standardization, collection. Only then can we store and analyze location not

as place per se, but as data. Amassing location data lets firms detect traffic jams without needing to see the cars: the number and speed of phones traveling on a highway reveal this information.

**Example of Big Data and insurances:** In the U.S. and Britain, drivers can buy car insurance priced according to where and when they actually drive, not just pay an annual rate based on their age, sex, and past record. This approach to insurance pricing creates incentives for good behavior. It shifts the very nature of insurance from one based on pooled risk to something based on individual action. Tracking individuals by vehicles also changes the nature of fixed costs, like roads and other infrastructure, by tying the use of those resources to drivers and others who “consume” them.

**Reality Mining:** This refers to processing huge amounts of data from mobile phones to make inferences and predictions about human behavior. In one study, analyzing movements and call patterns allowed them to successfully identify people who had contracted the flu before they themselves knew they were ill.

**Datafication and social media:** The idea of Datafication is the backbone of many of the Web’s social media companies. Social networking platforms don’t simply offer us a way to find and stay in touch with friends and colleagues, they take intangible elements of our everyday life and transform them into data that can be used to do new things.

**Datafied mood:** In one study, reported in Science in 2011, an analysis of 509 million tweets over two years from 2.4 million people in 84 countries showed that people’s moods followed similar daily and weekly patterns across cultures around the world—something that had not been possible to spot before. Moods have been datafied. Datafication is not just about rendering attitudes and sentiments into an analyzable form, but human behavior as well.

**Measuring body data:** Another company, Basis, lets wearers of its wristband monitor their vital signs, including heart rate and skin conductance, which are measures of stress. Getting the data is becoming easier and less intrusive than ever. In 2009 Apple was granted a patent for collecting data on blood oxygenation, heart rate, and body temperature through its audio ear buds.

**Reusable data:** We’re capturing information and putting it into data form that allows it to be reused. This can happen almost everywhere and to nearly everything. GreenGoose, a startup in San Francisco, sells tiny sensors that detect motion, which can be placed on objects to track how much they are used. Putting it on a pack of dental floss, a watering can, or a box of cat litter makes it possible to datafy dental hygiene and the care of plants and pets.

**The internet of things:** The enthusiasm over the “internet of things”—embedding chips, sensors, and communications modules into everyday objects—is partly about networking but just as much about datafying all that surrounds us. Once the world has been datafied, the

More book summaries at [www.kimhartman.se](http://www.kimhartman.se)

Contact me at [Kim.hartman@hyperisland.se](mailto:Kim.hartman@hyperisland.se)

potential uses of the information are basically limited only by one's ingenuity. Maury datafied seafarers' previous journeys through painstaking manual tabulation, and thereby unlocked extraordinary insights and value. Today we have the tools (statistics and algorithms) and the necessary equipment (digital processors and storage) to perform similar tasks much faster, at scale, and in many different contexts. In the age of big data, even backsides have upsides

**Datafication and society:** Like those other infrastructural advances, it will bring about fundamental changes to society. Aqueducts made possible the growth of cities; the printing press facilitated the Enlightenment, and newspapers enabled the rise of the nation state. But these infrastructures were focused on flows—of water, of knowledge. So were the telephone and the Internet. In contrast, Datafication represents an essential enrichment in human comprehension. With the help of big data, we will no longer regard our world as a string of happenings that we explain as natural or social phenomena, but as a universe comprised essentially of information. For well over a century, physicists have suggested that this is the case—that not atoms but information is the basis of all that is. This, admittedly, may sound esoteric. Through Datafication, however, in many instances we can now capture and calculate at a much more com.

**Big-data consciousness:** the presumption that there is a quantitative component to all that we do, and that data is indispensable for society to learn from.

## Chapter 6: Value

**Captcha:** The data had a primary use—to prove the user was human—but it also had a secondary purpose: to decipher unclear words in digitized texts.

**Data's value:** In the digital age, data shed its role of supporting transactions and often became the good itself that was traded. In a big-data world, things change again. Data's value shifts from its primary use to its potential future uses. In the age of big data, all data will be regarded as valuable, in and of itself. When we say "all data," we mean even the rawest, most seemingly mundane bits of information.

**Data have become accessible:** What makes our era different is that many of the inherent limitations on the collection of data no longer exist. Technology has reached a point where vast amounts of information often can be captured and recorded cheaply. Data can frequently be collected passively, without much effort or even awareness on the part of those being recorded. And because the cost of storage has fallen so much, it is easier to justify keeping data than discarding it. All this makes much more data available at lower cost than ever before.

**Data as resource:** In light of informational firms like Farecast or Google—where raw facts go in at one end of a digital assembly line and processed information comes out at the other—data is starting to look like a new resource or factor of production.

**Data is non-rivalrous:** Data's value does not diminish when it is used; it can be processed again and again. Information is what economists call a "non-rivalrous" good: one person's use of it does not impede another's. And information doesn't wear out with use the way material goods do.

**Data contains secondary value/Option value:** Just as data can be used many times for the same purpose, more importantly, it can be harnessed for multiple purposes as well. Data's full value is much greater than the value extracted from its first use. It also means that companies can exploit data effectively even if the first or each subsequent use only brings a tiny amount of value, so long as they utilize the data many times over. Data's true value is like an iceberg floating in the ocean. Only a tiny part of it is visible at first sight, while much of it is hidden beneath the surface. In short, data's value needs to be considered in terms of all the possible ways it can be employed in the future, not simply how it is used in the present

**Analogy between data and energy:** It may be helpful to envision data the way physicists see energy. They refer to "stored" or "potential" energy that exists within an object but lies dormant. Think of a compressed spring or a ball resting at the top of a hill. The energy in these objects remains latent—potential—until it's unleashed, say, when the spring is released or the ball is nudged so that it rolls downhill. Now these objects' energy has become

“kinetic” because they’re moving and exerting force on other objects in the world. After its primary use, data’s value still exists, but lies dormant, storing its potential like the spring or the ball, until the data is applied to a secondary use and its power is released anew. In a big-data age, we finally have the mindset, ingenuity, and tools to tap data’s hidden value.

**Option value:** The crux of data’s worth is its seemingly unlimited potential for reuse: its option value. Collecting the information is crucial but not enough, since most of data’s value lies in its use, not its mere possession. There are three potent ways to unleash data’s option value:

- basic reuse
- merging datasets
- finding “twofers”

**The reuse of data:** A classic example of data’s innovative reuse is search terms. At first glance, the information seems worthless after its primary purpose has been fulfilled. Companies like Hitwise, a web-traffic-measurement company owned by the data broker Experian, lets clients mine search traffic to learn about consumer preference.

**Recombinant data:** Sometimes the dormant value can only be unleashed by combining one dataset with another, perhaps a very different one. With big data, the sum is more valuable than its parts, and when we recombine the sums of multiple datasets together, that sum too is worth more than its individual ingredients. Today Internet users are familiar with basic “mashups,” which combine two or more data sources in a novel way.

**Extensible data:** One way to enable the reuse of data is to design extensibility into it from the outset so that it is suitable for multiple uses. For instance, some retailers are positioning store surveillance cameras so that they not only spot shoplifters but can also track the flow of customers through the store and where they stop to look. The extra cost of collecting multiple streams or many more data points in each stream is often low. So it makes sense to gather as much data as possible, as well as to make it extensible by considering potential secondary uses at the outset. That increases the data’s option value. The point is to look for “twofers” — where a single dataset can be used in multiple instances if it can be collected in a certain way. Thus the data can do double duty.

**Depreciating value of data:** Most data loses some of its utility over time. In such circumstances, continuing to rely on old data doesn’t just fail to add value; it actually destroys the value of fresher data. So the company has a huge incentive to use data only so long as it remains productive. It needs to continuously groom its troves and cull the information that has lost value. The challenge is knowing what data is no longer useful. Just basing that decision on time is rarely adequate.

More book summaries at [www.kimhartman.se](http://www.kimhartman.se)

Contact me at [Kim.hartman@hyperisland.se](mailto:Kim.hartman@hyperisland.se)

**The value of data exhaust:** A term of art has emerged to describe the digital trail that people leave in their wake: “data exhaust.” It refers to data that is shed as a byproduct of people’s actions and movements in the world. For the Internet, it describes users’ online interactions: where they click, how long they look at a page, where the mouse-cursor hovers, what they type, and more. Many companies design their systems so that they can harvest data exhaust and recycle it, to improve an existing service or to develop new ones. Google is the undisputed leader. It applies the principle of recursively “learning from the data” to many of its services. Every action a user performs is considered a signal to be analyzed and fed back into the system. Data exhaust is the mechanism behind many services like voice recognition, spam filters, language translation, and much more. When users indicate to a voice-recognition program that it has misunderstood what they said, they in effect “train” the system to get better.

**The discrimination of data in corporation valuing:** there is widespread agreement that the current method of determining corporate worth, by looking at a company’s “book value” (that is, mostly, the worth of its cash and physical assets), no longer adequately reflects the true value. The difference between a company’s book value and its market value is accounted for as “intangible assets. Intangible assets are considered to include brand, talent, and strategy—anything that’s not physical and part of the formal financial-accounting system. There is currently no obvious way to value data. The day Facebook’s shares opened, the gap between its formal assets and its unrecorded intangible value was nearly \$100 billion.

## Chapter 7: Implications

**Data, skills and ideas:** The types of big-data companies have cropped up, which can be differentiated by the value they offer. Think of it as **the data, the skills, and the ideas**.

1. **First is the data.** These are the companies that have the data or at the least have access to it. But perhaps that is not what they are in the business for. Or, they don't necessarily have the right skills to extract its value or to generate creative ideas about what is worth unleashing. The best example is Twitter, which obviously enjoys a massive stream of data flowing through its servers but turned to two independent firms to license it to others to use.
2. **Second are skills.** They are often the consultancies, technology vendors, and analytics providers who have special expertise and do the work, but probably do not have the data themselves nor the ingenuity to come up with the most innovative uses for it. In the case of Walmart and Pop-Tarts, for example, the retailer turned to the specialists at Teradata, a data-analytics firm, to help tease out the insights.
3. **Third is the big-data mindset.** For certain firms, the data and the know-how are not the main reasons for their success. What sets them apart is that their founders and employees have unique ideas about ways to tap data to unlock new forms of value. An example is Pete Warden, the geeky co-founder of Jetpac, which makes travel recommendations based on the photos users upload to the site.

**Banks and data:** The larger banks and the card issuers like Visa and MasterCard seem to be in the sweet spot of the information value chain. By serving many banks and merchants, they can see more transactions over their networks and use them to make inferences about consumer behavior. Their business model shifts from simply processing payments to collecting data. It discovered, among other things, that if people fill up their gas tanks at around four o'clock in the afternoon, they're quite likely to spend between \$35 and \$50 in the next hour at a grocery store or restaurant. A marketer might use that insight to print out coupons for a nearby supermarket on the back of gas-station receipts around that time of day.

**Data specialists:** The second category consists of data specialists- companies with the expertise or technologies to carry out complex analysis.

**Big-data mindset:** The third group is made up of companies and individuals with a big-data mindset. Their strength is that they see opportunities before others do—even if they lack the data or the skills to act upon those opportunities. The entrepreneurs with the big-data mindset often don't have the data when they start. But because of this, they also don't have

the vested interests or financial disincentives that might prevent them from unleashing their ideas.

**Microchips in cars:** Cars today are stuffed with chips, sensors, and software that upload performance data to the carmakers' computers when the vehicle is serviced. Typical mid-tier vehicles now have some 40 microprocessors; all of a car's electronics account for one-third of its costs. This makes the cars fitting successors to the ships Maury called "floating observatories." The ability to gather data about how car parts are actually used on the road—and to reincorporate this data to improve them—is turning out to be a big competitive advantage for the firms that can get hold of the information.

Today, in big data's early stages, the ideas and the skills seem to hold the greatest worth. But eventually most value will be in the data itself. This is because we'll be able to do more with the information, and also because data holders will better appreciate the potential value of the asset they possess.

**Future vision:** The biggest impact of big data will be that data-driven decisions are poised to augment or overrule human judgment. The subject-area expert, the substantive specialist, will lose some of his or her luster compared with the statistician and data analyst, who are unfettered by the old ways of doing things and let the data speak. This means that the skills necessary to succeed in the workplace are changing. To be sure, subject-area experts won't die out. But their supremacy will ebb. From now on, they must share the podium with the big-data geeks, just as princely causation must share the limelight with humble correlation. This transforms the way we value knowledge, because we tend to think that people with deep specialization are worth more than generalists—that fortune favors depth. Yet expertise is like exactitude: appropriate for a small-data world where one never has enough information, or the right information, and thus has to rely on intuition and experience to guide one's way. In such a world, experience plays a critical role, since it is the long accumulation of latent knowledge—knowledge that one can't transmit easily or learn from a book, or perhaps even be consciously aware of—that enables one to make smarter decisions.

**Big Data in gaming:** On the surface, online gaming allows Zynga to look at usage data and modify the games on the basis of how they're actually played. So if players are having difficulty advancing from one level to another, or tend to leave at a certain moment because the action loses its pace, Zynga can spot those problems in the data and remedy them. But what is less evident is that the company can tailor games to the traits of individual players. There is not one version of Farmville—there are hundreds of them. Zynga's big-data analysts study whether sales of virtual goods are affected by their color, or by players' seeing their friends using them

**Scale matters:** Scale still matters, but it has shifted. What counts is scale in data. This means holding large pools of data and being able to capture ever more of it with ease. Thus large

More book summaries at [www.kimhartman.se](http://www.kimhartman.se)

Contact me at [Kim.hartman@hyperisland.se](mailto:Kim.hartman@hyperisland.se)

data holders will flourish as they gather and store more of the raw material of their business, which they can reuse to create additional value.

**No medium way:** In traditional sectors, medium-sized firms exist because they combine a certain minimum size to reap the benefits of scale with a certain flexibility that large players lack. But in a big-data world, there is no minimum scale that a company must reach to pay for its investments in production infrastructure. Big data squeezes the middle of an industry, pushing firms to be very large, or small and quick, or dead.

More book summaries at [www.kimhartman.se](http://www.kimhartman.se)

Contact me at [Kim.hartman@hyperisland.se](mailto:Kim.hartman@hyperisland.se)

## **Chapter 8-9: Risks/Control**

These chapters would probably be interesting but isn't relevant for the purpose that I'm reading this book. Therefore, no notes.

## Chapter 10: Next

**Comments against causation:** I am not interested in causation except as it speaks to action,” explains Flowers. “Causation is for other people, and frankly it is very dicey when you start talking about causation. I don’t think there is any cause whatsoever between the day that someone files a foreclosure proceeding against a property and whether or not that place has a historic risk for a structural fire. I think it would be obtuse to think so. And nobody would actually come out and say that. They’d think, no, it’s the underlying factors. But I don’t want to even get into that. I need a specific data point that I have access to, and tell me its significance. If it’s significant, then we’ll act on it. If not, then we won’t.

**Human’s quest for understanding:** A worldview we thought was made of causes is being challenged by a preponderance of correlations. The possession of knowledge, which once meant an understanding of the past, is coming to mean an ability to predict the future. The idea that our quest to understand causes may be overrated—that in many cases it may be more advantageous to eschew why in favor of what—suggests that the matters are fundamental to our society and our existence.

**Data takes center stage:** Ultimately, big data marks the moment when the “information society” finally fulfills the promise implied by its name. The data takes center stage. All those digital bits that we have gathered can now be harnessed in novel ways to serve new purposes and unlock new forms of value

**The world of information:** We can capture and analyze more information than ever before. The scarcity of data is no longer the characteristic that defines our efforts to interpret the world. We can harness vastly more data and in some instances, get close to all of it. But doing so forces us to operate in untraditional ways and, in particular, changes our idea of what constitutes useful information. Instead of obsessing about the accuracy, exactitude, cleanliness, and rigor of the data, we can let some slack creep in.

**What instead of why:** Because correlations can be found far faster and cheaper than causation, they’re often preferable. For many everyday needs, knowing what not why is good enough. And big-data correlations can point the way toward promising areas in which to explore causal relationships.

**More data:** While the tools are important, a more fundamental reason is that we have more data, since more aspects of the world are being datified.

**Option value:** Much of the value of data will come from its secondary uses, its option value, not simply its primary use

**Data exhaust:** Sometimes an important asset will not be just the plainly visible information but the data exhaust created by people's interactions with information, which a clever company can use to improve an existing service or launch an entirely new one.

**History:** As big data becomes commonplace, it may well affect how we think about the future. Around five hundred years ago, humanity went through a profound shift in its perception of time, as part of the move toward a more secular, science-based, and enlightened Europe. Before that, time was experienced as cyclical, and so was life. Every day (and year) was much like the one before, and even the end of life resembled its start, as adults again became childlike. Later, time came to be seen as linear—an unfolding sequence of days in which the world could be shaped and life's trajectory influenced. If earlier, the past, present, and future had all been fused together, now humanity had a past to look back upon, and a future to look forward to, as it shaped its present. One of the defining features of modern times is our sense of ourselves as masters of our fate; this attitude sets us apart from our ancestors, for whom determinism of some form was the norm. Yet big-data predictions render the future less open and untouched.

**Big data predictions:** Nothing is preordained, because we can always respond and react to the information we receive. Big data's predictions are not set in stone—they are only likely outcomes.

**Messiness:** Messiness is an essential property of both the world and our minds; in both cases, we only benefit by accepting it and applying it.

**Big Data and innovation:** Big data enables us to experiment faster and explore more leads. These advantages should produce more innovation. But the spark of invention becomes what the data does not say. That is something that no amount of data can ever confirm or corroborate, since it has yet to exist. If Henry Ford had queried big-data algorithms for what his customers wanted, they would have replied "a faster horse" (to rephrase his famous saying). In a world of big data, it is our most human traits that will need to be fostered—our creativity, intuition, and intellectual ambition—since our ingenuity is the source of our progress.

More book summaries at [www.kimhartman.se](http://www.kimhartman.se)

Contact me at [Kim.hartman@hyperisland.se](mailto:Kim.hartman@hyperisland.se)

## Previous book summaries

[Virus of the mind by Richard Brodie](#)

[Connected by Nicholas Christakis and James Fowler](#)

[The Power of Habit by Charles Duhigg](#)

[Eating the Big Fish by Adam Morgan](#)

[Storytelling – Branding in practice by Klaus Fog](#)

[The Switch – How to change things when change is hard by Chip & Dan Heath](#)

[A Whole New Mind: Why Right-Brainers Will Rule the Future](#)

[The Element – How finding your passion changes everything by Ken Robinson](#)

[Disciplined Dreaming: A Proven System to Drive Breakthrough Creativity by Josh Linkner](#)

[Bounce – The myth of talent and the power of practice by Matthew Syed](#)

[The Two-Second Advantage by Vivek Ranadive and Kevin Maney](#)

[The Idea Writers by Teresa Iezzi](#)

[Velocity – The seven new laws of a world gone digital](#)

[Start With Why by Simon Sinek](#)