The nominal title of my talk is, “The Age of Big Data.”

I just lifted the title as a starting point here. It was the headline on a long piece I wrote back in February 2012 for the Sunday Review section of The New York Times. Headlines aren’t known for understatement, and this one certainly had a reaching quality to it – pushing the available evidence pretty hard and into the future.

But it also captured, in a few words, the view of many people across a spectrum of disciplines, not just technologists but intellectuals as well.

In the article, the money quote, as we say in journalism, came from Gary King, director of Harvard’s Institute for Quantitative Social Science. “It’s a revolution,” King said. “We’re really just getting under way. But the march of quantification, made possible by enormous new sources of data, will sweep through academia, business and government. There is no area that is going to be untouched.”

 Incidentally, that piece, published a year and a half ago, was generated the old-

fashioned way. A senior editor called me, and said, “Here’s the headline.

You figure out the story.”

In fact, it’s been years now that I’ve been trying to figure it out – that is, the implications of a cluster of advancing technologies that now fly under the banner of Big Data. So today, I thought I’d offer a few observations on the story so far – and try to put it in a historical context.

My wife mockingly calls it “Big Dadda, Big Dadda.” She’s mimicking Elizabeth Taylor’s clipped, vaguely frantic Southern accent in the classic movie, “Cat on a Hot Tin Roof.” Taylor plays Maggie the Cat. Her father-in-law, played by Burl Ives, is “Big Daddy” -- the overbearing patriarch of a wealthy family in a small Southern town.

 So what is Big Data? It’s a meme and a marketing term for sure.

But substantively, I think of it as three things. First, it is a bundle of technologies. Second, it is a potential revolution in measurement. And third, it is a point of view, or philosophy, about how decisions will be – and perhaps should be – made in the future.

The bundle of technologies, of course, is partly all the old and new sources of data – Web pages, browsing habits, sensor signals, social media messages, GPS data, genomic information and so on. As you know, it is also a set of enabling technologies: Hadoop, NoSQL, in-memory processing, the cloud computing model, and so on. The next technological building block is a higher-level ingredient – clever software tools largely taken from the steadily evolving world of artificial intelligence, notably machine learning.

 Hal Varian, chief economist of Google and an emeritus professor at Berkeley, has written a new paper for the Journal of Economic Perspectives, titled, “Big Data: New Tricks for Econometrics.” In the paper, he writes, “My standard advice to graduate students these days is to go to the computer science department and take a class in machine learning.”

Big Data, its proponents insist, is also shifting the center of gravity in decision-making. Decisions of all kinds, we’re told, will be made increasingly based on data and analysis rather than experience and intuition. That is the broader dimension of this Big Data phenomenon. And it is what David Brooks, my colleague and an Op-ed columnist at the Times, meant when he began a column earlier this year with this sentence: “If you asked me to describe the rising philosophy of the day, I’d say it is data-ism.”

But let’s step back for a little historical perspective. The growth of information – or data – has almost always been a challenge in its time – and that has often led to the development of new technologies.

 In the late 19th century, for example, the population of the United States was surging, and the Census Bureau couldn’t keep up. The 1880 census took eight years to complete the counting and tabulations. By 1890, the American population grew to 63 million – the Big Data of its day. But the 1890 population total was counted within a few weeks, and the answers to all the census questions were sorted and tablulated within a year.

(PIX – DATA SCIENTIST CIRCA 1890)

The difference? A cutting-edge technology had been deployed for the 1890 census. Punched cards. Developed by Herman Hollerith.

And those Hollerith punched cards would be the founding technology of the company that became IBM.

In the discussion of Big Data, there is a lot of focus on the sheer

 volume of data and it exponential growth path. I’m sure you’re familiar with the trends here. The volume of data is estimated to be doubling every two years. And researchers at Berkeley did the neat calculation that two days of current global data production – five quintillion bytes –is about equal to all the world’s conversations, ever. But the overall numbers, of course, are inflated by the data-density of modern video capture, production and distribution on YouTube and elsewhere. So the value of data volume is easily overstated. There’s a lot of water in the ocean too. But you can’t drink it.

But it is the increasing variety of data – from smartphones to sensors – far more than the volume that may well open the door to what some people call a revolution in measurement. This technology, it seems, could be the digital equivalent of the telescope or the microscope. Both those made it possible to see and measure things as never before – with the telescope, it was the heavens and new galaxies; and with the microscope, it was the mysteries of life down to the cellular level.

(PIX – A MEASUREMENT REVOLUTION)

 This guy over here is Antonie Van Leeuwenhoek, a 17th century scientist. He invented several microscopes, and is considered the “Father of Microbiology.”

In the Big Data era, much of what can be measured in fine detail is human behavior – what people are doing in the physical world or looking for online. We’ve seen early examples of the predictive power of this kind of measurement.

The best-known examples are probably mining the trends in Google search queries to predict changes in the offline world. In one continuing research project, begun in 2009 and updated repeatedly, Lynn Wu of the Wharton School and Erik Brynjolfsson of MIT tracked the frequency of search terms like “house prices, “real estate agent” and “mortgage rates” – and modeled the correlation between housing-related search terms and house sales The higher the frequency of those searches, the more likely the national housing market would heat up, and vice versa. In the most recent version, their model using search data predicted future home sales 24 percent more accurately than forecasts by experts from the National Association of Realtors.

But we’ve also seen the shortcomings of this approach of prediction by data correlation. The most prominent example was the performance of Google Flu Trends in the past flu season. Based on flu-related search terms, Google Flu Trends estimated that nearly 11 percent of Americans were ill at the January peak. That was nearly twice the level that actually occurred based on doctors’ reports to the Centers for Disease Control and Prevention. Apparently, Google’s algorithms were unable to sift out the effect of news reports and social media messages warning of a harsh flu season, which sent flu-related searches spiking.

Sometimes, correlation isn’t enough. For correlation alone can leave out things – context, in his case – that shapes outcomes. Correlation does not address the “why” questions – why things happen.

Still, data-related correlations can often be powerful, illuminating or even amusing. For example, researchers at the MIT Media Lab, using cellphone GPS tracking and other data, have been able to predict if a person is a good credit risk more accurately than the official FICO scores of creditworthiness. And Place IQ, a smartphone location analysis firm, has determined that the three times people are most likely to click on a mobile ad are: When they are in a theater, waiting for a movie to start; when they are fishing; and on Sunday mornings.

 We’re just on the cusp of this revolution in measurement. You sense, you measure, you communicate – and your can change behavior to improve health and wellness, save energy, prevent crime, manage traffic, conserve water. Start-ups, big companies, governments and universities are working on all those.

Steven Koonin, a long-time physicist at CalTech and former undersecretary for science in the Energy Department in the Obama administration, plans to create an “an urban observatory,” as he puts it. The city whose life is to measured as never before is New York, and Koonin is the director of NYU’s Center for Urban Science and Progress. His new center intends to use all the publicly available government data in the city and combine that with thermal, imaging and even chemical sensor data.

Koonin said, “It feels these days as it must have felt when Galileo first turned his telescope to the heavens or when Van Leeuwenhoek first looked at a living cell.”

“What we can do today with observation and measurement,” he added, “is qualitatively different.”

(PIX – WAY BEYOND ORWELLIAN)

Obviously, this kind of measuring and monitoring – every clickstream, every movement, every commercial transaction – raises far-reaching questions about privacy and surveillance. When can companies and governments collect data? And when can they use it? The uproar surrounding the disclosure of the National Security Agency’s programs to mine telephone call records and e-mail message traffic is just the most recent round in that debate. It’s an important subject, and a good one. But it’s also a whole other discussion – and a detailed one –about policy and regulation, social values and trade-offs involving privacy, security and commerce. And it’s not one I’m going to take up today, other than mentioning it in passing.

(PIX – THE MEASUREMENT PARADOX)

Instead, I wanted to make a couple of observations on Big Data and the future of decision-making. This technology may well affect everything from how companies are run to how people make decisions in their own lives.

To frame the issue, two famous quotes come to mind.

 The first is: “You can’t manage what you can’t measure.”

 For this one, there seem to be twin claims for attribution – either W. Edwards Deming, statistician and quality control expert, or Peter Drucker, management consultant.

Who said it first doesn’t really matter. It’s a mantra in business and has the ring of commonsense truth.

 But there’s a lot of truth in the next one too.

“Not everything that can be counted counts, and not everything that counts can be counted.”

 This one is often attributed to Albert Einstein, but the stronger claim of origin probably goes to the sociologist William Bruce Cameron. But again, who said it first matters far less than what it says.

 Big Data can be seen as the next step in management by measurement. These technologies are tools that are here and will be used. That’s a good thing, in general. But I’d suggest that the enthusiasm for the kind of monitoring and measurement that Big Data makes possible should be balanced with a dose of the humility found in that second quote.

 So yes, data-driven decision-making will be the wave of the future. Data, for example, is an antidote to the human tendency to rely too much on a single piece of information or what is familiar – what psychologists call “anchoring bias.” Data can combat the bias of the familiar and suggest alternative answers that otherwise might be missed. It promises to make decisions of all kinds more scientific.

 (PIX – SCIENTIFIC MANAGEMENT)

 Still, there is a case for caution. Big Data is a descendant of Frederick Winslow Taylor’s “scientific management” of a century ago. Taylor’s instruments of measurement and recording were the stopwatch, clip board and his own eyes. Taylor and his acolytes used these time-and-motion studies to redesign work for maximum efficiency.

 Yet eventually, the excesses of that approach became apparent – and even satirical grist for Charlie Chaplin’s “Modern Times.” It might seem easy to dismiss Taylorism as a simple-minded throwback to another era. But scientific management was seen by many in its day a modernizing, progressive movement -- a way to rationalize work for the benefit of both workers and management.

 In his “Principles of Scientific Management,” published in 1911, Taylor made the kind of broad claims for his tools of measurement, observation and scientific decision-making that we hear today about Big Data. The principles of scientific management, Taylor wrote, “can be applied with equal force to all social activities; to the management of our homes; the management of our farms; the management of the business of our tradesmen, large and small; of our churches, our philanthropic institutions, of our universities, and our government departments.”

 Ever since Taylor’s scientific management, the enthusiasm for quantitative methods has waxed and waned in cycles.

(PIX – MODELS BEHAVING BADLY)

 The software algorithms of Big Data that mine vast stores of digital information, looking for patterns, correlations and anomalies, can trace their heritage partly to finance. Wall Street’s computer and math wizards are known as “quants.” And on Wall Street, we see both the power and the peril of the Big Data approach.

 Computerized math models didn’t cause the financial crisis, but they certainly played a role. Shortly after the crisis hit in 2008, I talked to Emanuel Derman – one of the original quants at Goldman Sachs. He’s now a professor at Columbia University. The trouble, he explained, comes when the predictive models attach crisp, firm numbers to messy, often unpredictable human behavior – say, the probability that a person will pay back a mortgage when the economy sours or the viral spread of panic in a market crash.

 The data models have their uses. But as Derman said, “Anyone who mistakes the model for the real world is a fool.”

 (PIX – IN PRAISE OF INTUITION)

 Decisions should be made based on data and analysis rather than experience and intuition. More science and less gut feel. Who could possibly argue with that? But let me suggest a caveat. At its best, what we call experience and intuition is really the synthesis of vast amounts of data, but the kind of data that we can’t attach reliable numbers to.

 This is what Steve Jobs called “taste.” He used to say that an enriched life involved seeking out and absorbing the best of your culture – whether in the arts or software design – and that would shape your view of the world and your decisions.

 I’ll tell just one Steve Jobs story on this theme. In early 2010. Apple’s iPad had been announced, but it was not yet on sale. Steve Jobs was at the New York Times, showing off the device to a dozen or so editors and reporters around the company’s boardroom table.

 An editor asked how much market research that had gone into the iPad. Those of us who knew Jobs pretty much knew what the answer would be. “None,” Jobs replied. “It’s not the consumers’ job to know what they want.”

 The counterargument is that Steve Jobs was a unique person. I take the point. But I do think there is a more general lesson – and so do others.

Sandy Pentland is a computational social scientist at MIT, and he believes that we will all soon be living in what he calls a “data-driven society.” He’s a believer. But as he puts it, Big Data technology is good at interpolation but not good at extrapolation. In short, Big Data fails when a decision requires an intuitive step outside the data sandbox – beyond the range of the data.

 No company has embraced Big Data more than IBM. Today, it has 9,000 consultants and 400 mathematicians working on Big Data projects. A push for all that came in November 2008, when IBM launched its Smarter Planet campaign. Sam Palmisano, who was the chief executive, introduced the concept in a speech at the Council on Foreign Relations. But his speech was made in the teeth of the financial crisis, when the global economy was in a tailspin. And yet he was pointing to the big opportunity on the horizon for the technologies now called Big Data. It could have been nothing more than a tone-deaf marketing pitch.

 But his prediction turned out to be on target. The Smarter Planet initiative proved to be a success, and a big business for IBM. Three years later, when he was stepping aside as chief executive, I asked Palmisano about that decision and its timing. He said he talked to IBM’s scientists, gathered data on technology and market trends. Then, he made the bet, based on data but also intuition. “It was a judgment call,” he explained. “If it didn’t take judgment, a computer could do it.”

Well, not yet, it seems.

So I see Big Data as what I think of as a yes-but story. Yes, this technology is coming. It’s unavoidable. But it’s a tool, a means to improve human decision-making – not replace it.

And that observation is merely an echo of the insight – and prediction – made back in 1960 by J.C.R Licklider, the renowned psychologist and computing pioneer. In his classic essay, “Man-Computer Symbiosis,” Licklider declared that the appropriate goal of computing was to, as he put it, “augment” human intelligence rather than substitute for it.