SPE 170633-MS

Data as an Asset: What the Upstream Oil & Gas Industry Can Learn About "Big Data" from Companies like Facebook

Robert K. Perrons, Queensland University of Technology
Jesse W. Jensen, Intel Corporation

This paper was prepared for presentation at the SPE Annual Technical Conference and Exhibition held in Amsterdam, The Netherlands, 27–29 October 2014.

Abstract

The upstream oil & gas industry has been contending with massive data sets and monolithic files for many years, but "Big Data"—that is, the ability to apply more sophisticated types of analytical tools to information in a way that extracts new insights or creates new forms of value—is a relatively new concept that has the potential to significantly re-shape the industry. Despite the impressive amount of value that is being realized by Big Data technologies in other parts of the marketplace, however, much of the data collected within the oil & gas sector tends to be discarded, ignored, or analyzed in a very cursory way. This paper examines existing data management practices in the upstream oil & gas industry, and compares them to practices and philosophies that have emerged in organizations that are leading the Big Data revolution. The comparison shows that, in companies that are leading the Big Data revolution, *data is regarded as a valuable asset.* The presented evidence also shows, however, that this is usually not true within the oil & gas industry insofar as data is frequently regarded there as descriptive information about a physical asset rather than something that is valuable in and of itself. The paper then discusses how upstream oil & gas companies could potentially extract more value from data, and concludes with a series of specific technical and management-related recommendations to this end.

1. Introduction

The concept of collecting large data sets and then analyzing them has been around for a long time. Developments in mathematics as far back as the 17th and 18th centuries gave rise to a range of new tools that could more precisely predict the movements of celestial bodies, and that could predict with reasonable accuracy the rates of crime, marriage, and suicide within a given population (White House, 2014). Society and the scientific community have steadily improved their ability to collect information and analyze it since that time, and data is now more engrained in our lives and economy than ever before. As a recent report to the U.S. President suggests, the "collection, storage, and analysis of data is on an upward and seemingly unbounded trajectory, fueled by increases in processing power, the cratering costs of computation and storage, and the growing number of sensor technologies embedded in devices of all kinds" (White House, 2014, p. 1).[1]

The exploration & production (E&P) sector has already been impacted by several of these trends. Modern seismic data centers

[1] The growing use of relatively inexpensive sensors such as radio frequency identification (RFID) tags or accelerometers in day-to-day devices that are linked through wired and wireless networks has been labeled as the "Internet of things" (White House, 2014). These devices typically use the Internet to transmit, compile, and analyze the data, and their proliferation—and the vast amounts of digital information that they generate—has been a significant contributing factor to the Big Data revolution.

like those from the oil & gas sector can easily contain as much as 20 petabytes[2] of information, which is equivalent to 926 times the size of the U.S. Library of Congress. If this amount of information was copied into books and put on a single continuous bookshelf, it would go around the Earth's equator approximately six times (Beckwith, 2011). And while seismic data sets are notoriously large and cumbersome, many other aspects of the E&P industry are also generating significantly more data than they used to. What is more, there's every reason to believe that this trend towards more digital information is just getting warmed up. Current estimates suggest that the total amount of digital data in the world—including things like books, images, e-mails, music, and video—is doubling every 2-3 years (Lohr, 2012a; Mayer-Schönberger & Cukier, 2013).

"Big Data"[3] is a rather vague term that describes the application of new tools and techniques to digital information on a size and scale well beyond what was possible with traditional approaches (Lohr, 2012b). The idea has grown organically from a broad range of academic disciplines, technology providers, and consultants and, as a consequence, no single definition has been agreed for this steadily evolving research area. Despite the absence of a precise definition, however, most users of the term broadly agree that it refers to data sets and analytical techniques in applications that are so large and complex that they require advanced data storage, management, analysis, and visualization technologies (Chen, Chiang, & Storey, 2012). These new capabilities have most famously delivered value in the social media sector and retail industries (Harford, 2014; Mayer-Schönberger & Cukier, 2013), but have also led to major breakthroughs in a diverse range of other contexts, including scientific research (Frankel & Reid, 2008; Kluger, 2014; Lynch, 2008), healthcare (Chen, et al., 2012), the oil & gas industry (Anand, 2013; Feblowitz, 2013), heavy equipment manufacturers such as Caterpillar[4] (Mehta, 2013), and professional sports (Leahey, 2013).[5] In fact, the transformational potential of Big Data is such that Gary King, the director of Harvard University's Institute for Quantitative Social Science, estimates that there "is no area that is going to be untouched" by it (Lohr, 2012a).

2. What Has Made Big Data Possible?

Big Data is not a result of a single silver-bullet technology, but rather the coming together of several innovations and novel ideas in a highly complementary way. Four of these technological developments are particularly noteworthy:

1) A precipitous decline in data storage costs. As shown in Appendix A, the cost of storing digital information has been falling at an exponential rate for a long time. Many years ago, it was standard practice in many industries to discard significant collections of data when their initial use had passed, as there was a real economic expense associated with archiving the data afterward (e.g., Feblowitz, 2013). This is much less true today.

2) Continued growth in the processing speeds of computing devices. Moore's Law, which states that the number of transistors on integrated circuits doubles approximately every two years, has been continuing unabated since the 1970s (Appendix B). It therefore follows that the amount of computing power available in commercially available devices has been increasing at a similarly impressive rate.

3) Breakthroughs in relevant areas of mathematics. Whereas traditional data sets have historically been fairly structured, orderly, and static, digital information in the era of Big Data is frequently noisy, messy, raw, unstructured, and dynamic (Ouellette, 2013). Recent developments in mathematics—most notably, geometry—have significantly helped Big Data practitioners see through the messiness of these new data sets to find useful information and relationships. For example, Professor Gunnar Carlsson at Stanford University has made considerable progress with representing massive data sets as networks of geometrical nodes and edges so that the data can be rationalized using a suite of mathematical tools known as topological data analysis (TDA). Simply put, TDA is a way of getting

[2] This statistic is less impressive when you consider that Walmart, the US retail giant, collects more than 2.5 petabytes of data every hour from customer transactions (McAfee & Brynjolfsson, 2012).

[3] The term "Big Data" is believed to have been coined by the astronomy and genomics communities in the 2000s (Mayer-Schönberger & Cukier, 2013, p. 6), but the concept has been used more widely since then.

[4] Caterpillar is proud of its attempt to connect the company's equipment "to an intelligent network that can monitor the gear and provide useful reports on equipment repairs, operator usage patterns, and the like… A dealer could use the information to keep an eye out for upcoming repairs and stock inventories accordingly. And site managers could monitor a fleet to determine if, say, operators are needlessly idling equipment" (Mehta, 2013, p. 34).

[5] The 2011 movie *Moneyball*, starring Brad Pitt, chronicles how the low-budget Oakland A's started using historical performance data and arcane baseball statistics to spot undervalued players. Intensive data analysis has since become commonplace not only in baseball but also in other sports, including English soccer (Lohr, 2012a).

structured information out of unstructured data so that machine-learning algorithms can be applied to it (Ouellette, 2013).

4) The development of software platforms such as Google's MapReduce or its open-source rival Hadoop.[6] These tools make it possible to break large data sets into smaller chunks that can be delegated to several computing devices. The results of the calculations arising from each of the smaller chunks can then be re-integrated at the end of the process. This approach frequently uses cloud computing infrastructure as a platform for transferring these data chunks to different computing devices, and then bringing back the results.

3. How is Big Data Different from What Was Done Previously?

Despite the considerable hype that has been generated around the topic of Big Data, the overarching objective is something that businesses have been aspiring to for a long time: to make better decisions (Regalado, 2014). What is changing, however, are the specific mechanisms by which these decisions are made. First, Big Data differs from traditional approaches on account of the "three Vs": volume, velocity, and variety (McAfee & Brynjolfsson, 2012). The steadily decreasing costs associated with collecting and storing data have resulted in a fundamental shift in thinking about data quality and volume. Historically, data collection was predicated on sampling from a subset of an overall population, and trying to make the collected data from that sample as accurate as possible. By stark contrast, the move towards Big Data has led to a much greater tolerance for messiness and imprecision. This more relaxed approach to vagueness has been compensated, however, by much larger volumes of data. Underlying this change of philosophy is the belief that "more trumps better" (Mayer-Schönberger & Cukier, 2013, p. 33).

These much larger volumes of data are now able to move with ever-increasing velocities such that we can monitor a bewildering number of system variables in nearly real-time. Moreover, this data is coming from a wider variety of sources and in an increasingly broad array of formats. Some data is "born digital," meaning that it was created specifically for digital use by a computer or data processing system (e.g., e-mail, web browsing, GPS locations); other data is "born analog," meaning that it comes from the physical world, but can increasingly be converted into a digital format (e.g., voice or visual information captured by phones, cameras, or video recorders, or data collected from wearable devices). The rising capability of "data fusion" makes it more and more possible to bring together disparate sources of data to glean fresh insights that nobody predicted (White House, 2014).

4. How Big Data is Unfolding Differently in the E&P Sector

In light of this sweeping global trend, it is hard to imagine a future in which the E&P industry isn't collecting significantly more data than it does at the moment. With such big piles of digital information accumulating around us, it's easy to understand why many in the E&P sector believe that we're solidly on track to reap the benefits of the "Big Data" revolution.

But the E&P sector seems to be approaching these rapidly growing piles of data with the same attitudes and analytical techniques that have been with us for years. As Feblowitz (2013) suggests, a lot of potentially valuable digital information harvested from upstream oil & gas assets is barely given a cursory glance, and much of it is simply thrown away. Moreover, in those instances where data is stored, it is often kept by the service companies responsible for generating it rather than the operator in charge of managing the long-term welfare of the asset.

This is not how the Big Data revolution is unfolding in many other industries. Several other sectors—most notably, the healthcare, financial, retail, and media industries—have come to realize that new and valuable insights are frequently gleaned from using new techniques to analyze massive data sets in ways that were never possible with smaller ones. These insights tend not to be discovered by testing hypotheses between variables whose relationships are well understood; rather, they are found by applying advanced analytical techniques to massive numbers of variables that, at first blush, might seem to be unrelated. FICO, an American analytics company, has discovered a surprisingly tight relationship between aspects of a person's car ownership records and their propensity to take prescribed medication. Aviva, a large UK-based insurance firm,

[6] Using Hadoop, Visa was able to reduce the processing time for two years' worth of test records—which translates to approximately 73 billion transactions—from one month to only 13 minutes (Mayer-Schönberger & Cukier, 2013, p. 46).

has developed predictive models that use credit reports and consumer-marketing data to identify health risks among prospective applicants. And by carefully examining what its customers are purchasing from different departments, Target, the US retailer, "knows when a woman is pregnant without the mother-to-be explicitly telling it so" (Mayer-Schönberger & Cukier, 2013, p. 57).

Companies that are leading the charge in the Big Data revolution are not merely creating value by monitoring relationships that they already knew about, but by finding patterns and making predictions based on complex relationships that were previously unknown. And whereas scientists have traditionally sought to understand the causality and mechanisms underlying these kinds of relationships, cutting-edge users of Big Data frequently care about the "what" far more than the "why." In other words, "When we let the data speak, we can make connections that we had never thought existed" (Mayer-Schönberger & Cukier, 2013, p. 14). This fundamentally different approach to data analysis carries with it a valuable lesson about the kinds of information that should be collected: in the age of Big Data, potential value is lurking within *all* digital information, no matter how inconsequential and disconnected it might seem at the time it is collected.

Therein lies a fundamental difference of opinion between the E&P sector and other industries that are considered leaders in Big Data. While there's no denying that the upstream oil & gas industry is swimming in digital information—and, indeed, several Big Data technologies have been used for much longer in this industry than in many others—the way that we manage data doesn't actually bear much resemblance to how it gets used in companies like Facebook and Amazon. The E&P sector tends to regard data as information that describes the state of an asset; leaders in Big Data, by stark contrast, realize that *data is a valuable asset in and of itself.*

It goes without saying that the business models behind social networks and online retailing are profoundly different from that of the E&P sector, and that data delivers value within the upstream oil & gas industry in a way that is unlike many other sectors. But at the same time, it's no less true that the competitive landscape of the E&P industry is growing ever more reliant on information technologies and computing power. The Internet began as a curious plaything at the beginning of that revolution but, over the past 20 years, it has clearly become an indispensable part of how we do business. And so it goes with Big Data: many of today's E&P companies clearly do not consider most of their digital information to be mission-critical to their profitability. But they probably will start to think of data that way in the years ahead.

In the future, the issue of data management—including how data will be collected, formatted, stored, and owned—will be an important part of service contract negotiations and agreements between collaborating firms. Companies within the industry will jealously guard the data that they hold and, as suggested by Anand (2013), they will strategically attempt to augment in-house data assets with digital information procured from outside sources, thereby giving them the kinds of massive data sets in which the value of Big Data is often lurking. And petroleum engineers in the years ahead will speak disparagingly about us because we failed to appreciate the long-term value of the data that we're casting aside so carelessly today.

5. Conclusions and Key Questions

This paper outlined the high-level capabilities and new technologies that have given rise to Big Data, and briefly examined the potential relevance of these changes to the E&P sector. The amount and kinds of value that Big Data can deliver will clearly vary considerably from one industry to the next, but a few important themes are emerging across the entire business landscape. Among the most important of these is an evolutionary change in the market's perceptions of data—in particular, the notion of data as an asset (Perrons & Jensen, 2014) described earlier. This fundamental shift raises several important questions for the E&P industry. Specifically:

1) In the future, significant strategic advantage will arise from an organization's ability to be the integrator and high-level analyzer of data within its ecosystem. How should E&P companies re-shape their contracting and collaboration strategies in light of this new reality?

2) Big Data is not merely about organizations doing what they have always been doing, but with more data. Instead, these forces are resulting in fundamentally new business models in many industries (Mayer-Schönberger & Cukier, 2013). So how could Big Data change the E&P industry's business model?

3) Finally, Big Data is resulting in noticeable changes in job descriptions and the internal architecture of organizations. A new breed of professionals known as "data scientists" has emerged that understands analytics, information technologies (IT), and mathematics while also having the ability to communicate effectively with decision-makers (Davenport, Barth, & Bean, 2012). This represents a marked departure from the data analysts that have been residing in many organizations for years. Should E&P companies be cultivating teams of data

scientists in order to address the rising importance and increasingly multi-disciplinary nature of IT and data management? A few universities and organizations are making inroads in this area (Mahdavi, 2008; Perrons, 2010) but, for the most part, the sector's discipline silos have remain largely unchanged.

The E&P sector's digital revolution is unfinished. The industry has succeeded handsomely in learning how to generate a staggering amount of data, but we're still collectively wrestling with the question of what to do with it. The case for moving towards digital oilfield technologies was largely based on the ability of those tools to help us make better decisions—and when you peel back all the hype, that is ultimately what Big Data is about, too (Regalado, 2014). In this way, Big Data isn't the dawn of a new age for the E&P sector, but rather the next phase of a digital transformation that started a long time ago. The industry's digital revolution will be complete when we come to terms with how to monetize the data that we're now capable of collecting and use it to create all the value that it can.

6. References

Anand, P. (2013). Guest Editorial: Big Data is a Big Deal. *Journal of Petroleum Technology, 65*(4), 18-21.

Anonymous. (2011). Moore's Law: The Economics of Abundance. Retrieved 15 May 2014 from http://letshavetheconversation.blogspot.com.au/2011/11/moores-law-economics-of-abundance.html

Beckwith, R. (2011). Managing Big Data: Cloud Computing and Co-Location Centers. *Journal of Petroleum Technology, 63*(10), 42-45.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly, 36*(4), 1165-1188.

Davenport, T. H., Barth, P., & Bean, R. (2012). How "Big Data" is Different. *MIT Sloan Management Review, 54*(1), 22-24.

Feblowitz, J. (2013). *Analytics in Oil and Gas: The Big Deal about Big Data*. Paper presented at the SPE Digital Energy Conference and Exhibition 2013 at The Woodlands, Texas, USA from 5-7 March 2013. Paper number SPE 163717.

Frankel, F., & Reid, R. (2008). Distilling Meaning from Data. *Nature, 455*(7209), 30.

Harford, T. (2014, March 28). Big Data: Are We Making a Big Mistake? *Financial Times*. Retrieved from http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2xV4TdTmn

Kluger, J. (2014). Finding a Second Earth. *Time, 183*(1), 30-32.

Komorowski, M. (2014). A History of Storage Costs. Retrieved 19 May 2014 from http://www.mkomo.com/cost-per-gigabyte

Leahey, C. (2013). The NFL's Plan to Tackle Big Data. *Fortune* (Asia-Pacific edition), *168*(9), 9.

Lohr, S. (2012a, February 11). The Age of Big Data. *New York Times*, p. Printed on p. SR1 of the New York edition of the newspaper. Retrieved from http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html

Lohr, S. (2012b, August 11). How Big Data Became So Big. *New York Times*, pp. Also published on p. BU3 of the August 12 issue of the New York Edition with the headline "Amid the Flood, a Catchphrase is Born.". Retrieved from http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html

Lynch, C. (2008). How Do Your Data Grow? *Nature, 455*(7209), 28-29.

Mahdavi, M. (2008). The Digital Petroleum Engineer: Carpe Diem! (Guest Editorial). *Journal of Petroleum Technology, 60*(10), 16-17.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London, United Kingdom: John Murray Publishers.

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review, 90*(10), 60-68.

Mehta, S. N. (2013). Where Brains Meet Brawn. *Fortune (Asia-Pacific edition), 168*(7), 34.

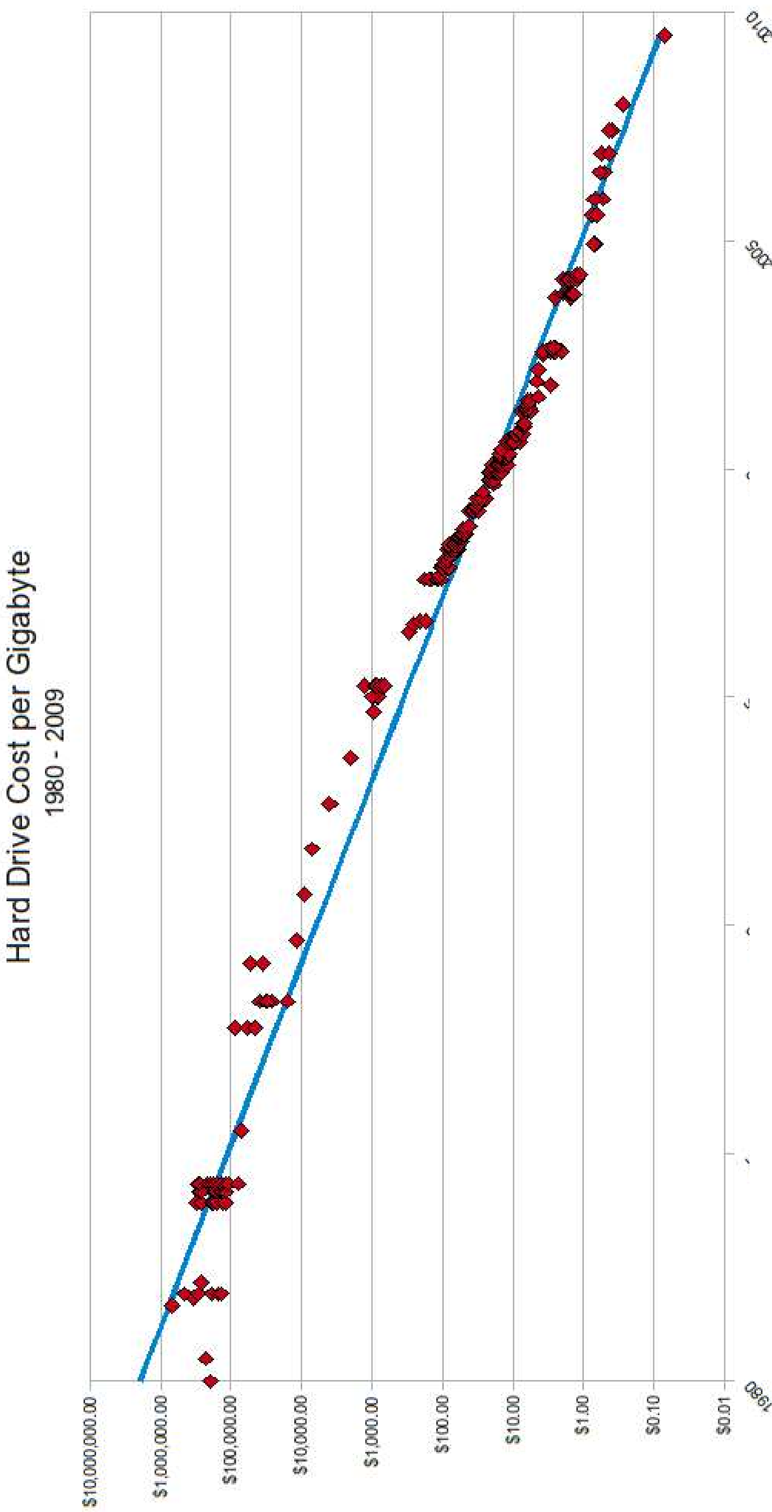
Ouellette, J. (2013, October 4). The Mathematical Shape of Things to Come. *Quanta Magazine* (online), from http://www.simonsfoundation.org/quanta/20131004-the-mathematical-shape-of-things-to-come/

Perrons, R. K. (2010). Organizational Flexibility in the Digital Age: The Stakes are Higher Than You Think. *Journal of Petroleum Technology, 62*(6), 16-18.

Perrons, R. K., & Jensen, J. (2014). The Unfinished Revolution: What is Missing from the E&P Industry's Move to "Big Data". *Journal of Petroleum Technology*, *66*(5), 20-22.

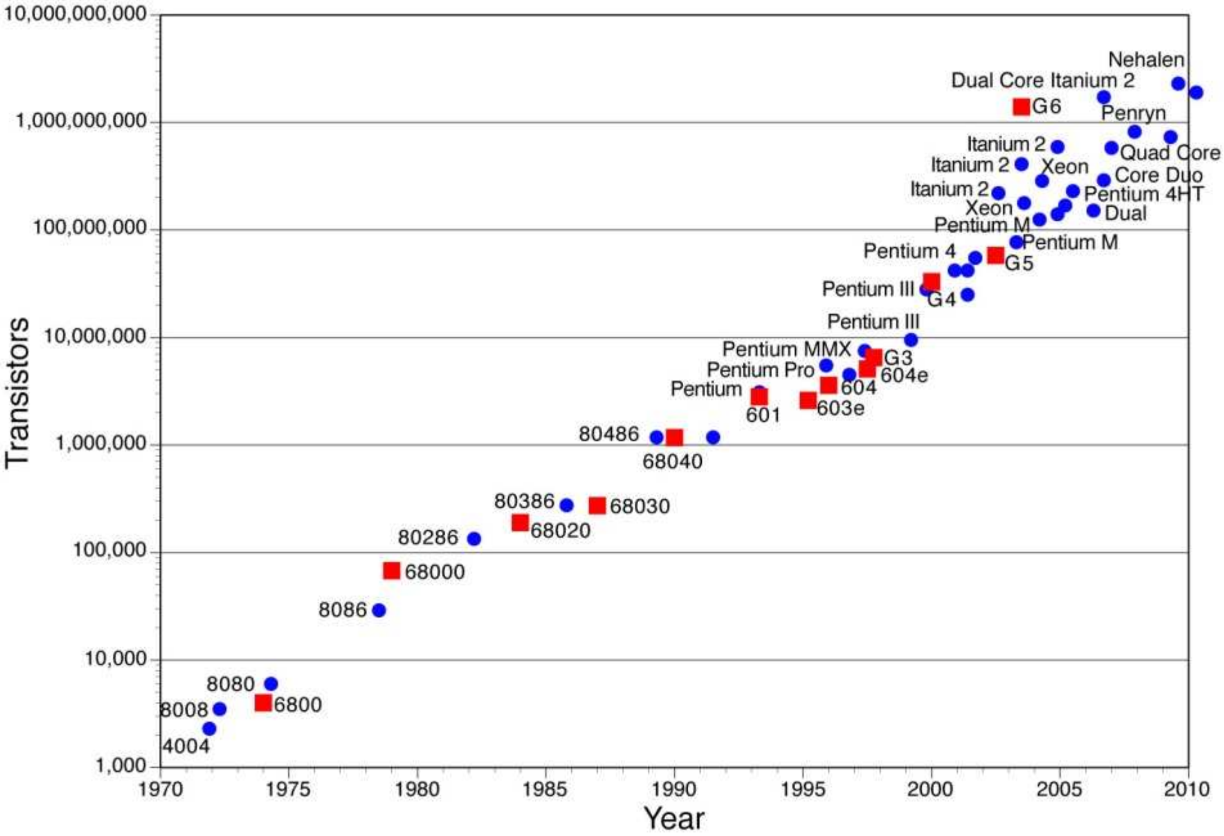
Regalado, A. (2014). Data and Decision Making. *MIT Technology Review, 117*(2), 61-63.

White House. (2014). *Big Data: Seizing Opportunities, Preserving Values*. Washington, DC: Executive Officer of the President.



Appendix A: Data Storage Costs, 1980-2009 (Komorowski, 2014)

Appendix B: Increase in Availability of Computing Power, 1970-2010



Moore's Law (Anonymous, 2011)