A Bug's Life: Investment opportunities in Big Data

From time to time, we focus on a specific industry or investment topic in the *Eye on the Market*. Recent issues covered distressed real estate, oil & gas investing, private credit "rescue" lending to troubled companies, the purchase of loan pools from over-leveraged European banks, debtor-in-possession financing and next-generation telecommunications. This week, some comments on "Big Data": what it means, and what investment opportunities it entails.

Big Data most often refers to amounts of information so large *and* variable in format that customized tools are required to store and analyze it. Why write about this? We are at a point in time where evolving technologies (cloud, social, mobile, etc) are changing the way people and corporations do everything from find a date to manage inventory. At the intersection of these trends sits enormous amounts of data being created, stored and analyzed in new ways. We can't cover all the changes in technology here; this is meant to take a look at one slice of a broader landscape. In this note, we focus on opportunities associated with companies that provide value-added database management tools and processing of Big Data, and companies that are transformed after using them. Note to Big Data experts and technology junkies: this paper is for laypeople, not clerics.

Big Data reminds me of the insect world: at first glance, it looks like an unmanaged swarm of chaos[1]. But after a closer look, you can see all the little bits and pieces being put to work in pursuit of an organized, holistic enterprise. To start, some simple mathematical conversions that show how much data gets created and stored as we do the things we take for granted.

Some simple data conversions, and how much storage is required in the digital world

Bit	Smallest unit of computer information: a binary yes/no indicator.
Byte	Equal to 8 bits. 100 bytes = a telegram, like the one JFK said he received from his father during the 1960 election which read: "*Dear Jack: Don't buy one more vote than necessary. I'll be damned if I pay for a landslide.*"
Kilobyte	1,000 bytes. 25 kilobytes = average email. 107 trillion emails were sent in 2010, 80%-90% of which were spam (*Pingdom*). 500 kilobytes = 10,000 computer punch cards. The Punch-card UNIVAC correctly predicted Eisenhower's landslide in 1952, but then its voting machine offspring wrecked the 2000 election in Florida.
Megabyte	1,000 kilobytes. A 3.5 inch diskette held 1.44 mb. 1 megabyte = 3 seconds of high definition 1080i60 video, perhaps of a water-skiing squirrel, or a guy putting a cell phone in a blender to see what happens (popular YouTube videos). 10 megabytes = 1 digital chest x-ray. 8 megabytes: *Remembrance of Things Pas*t, by Marcel Proust.
Gigabyte	1,000 megabytes. 1 gigabyte = 2x the data on a CD-ROM; = ten yards of books; = an iTunes movie in std definition.
Terabyte	1,000 gigabytes. 1 terabyte = all x-ray films in a large hospital; = 2,000 hours of music at CD quality. I have a 1 terabyte hard drive at home since I have 60 GB of music and my wife takes a lot of pictures of her relatives, Africa, furniture and fish I have caught (50 GB). July 2012 conquests include a 3.5 lb rainbow trout and 3 northern pike.
Petabyte	1,000 terabytes. 1 petabyte = 20 million filing cabinets of text. Wal-Mart's data warehouse holds 2.5 petabytes, equal to the information content of half the letters delivered by the US Postal Service in 2010. Astronomers expect to eventually process 10 petabytes of data per hour from the Square Kilometer Array telescope (CSIRO).
Exabyte	1,000 petabytes. Enterprises and individuals stored 13 exabytes of data in 2010. 1 exabyte = 4,000x the information stored in the US Library of Congress; 1 exabyte = 10,000 years of high definition 1080i60 video, or in terms of one single movie, 52 million viewings of "*Barbarella*", the first movie I remember seeing on cable television in 1980.

Before everyone's eyes glaze over, let's move to the next subject: where does all this data come from? Mostly, it comes from the fact that **the world has moved from "analog" to "digital".** In 1986, 99% of all information created was stored in books, pictures and on audio/video tapes; only 1% was stored digitally. By 1993, digital storage rose to 3%, and to 25% in 2000. Then, by 2007, 94% of all information created was stored in digital form. Here are a few snapshots of the digital avalanche:

A partial look at where all the data is coming from

- ***Mobile devices***. 4 billion people, or 60% of the world's population, use mobile phones. Of these, 12% are smartphones, a category growing at 20% per year. The average traffic per smartphone in 2011 was 150 MB per month, up from 55 MB per month in 2010. The rapidly growing use of tablets also generates location data and other user information. JP Morgan Securities LLC cites 70 million tablets sold in 2011, with shipments growing at 40% per year. According to Cisco, by the end of 2012, **the number of mobile-connected devices will exceed the number of people on earth**, and by 2016, there will be over 10 billion mobile-connected devices, exceeding the world's 7.3 billion population at that time[2].

[1] Insects are an apt metaphor given how many of them there are. According to the Smithsonian Institution, at any given point in time, there are 10 quintillion (10^19) insects alive on earth. In computer terms, that would be 10 exabytes of insects.

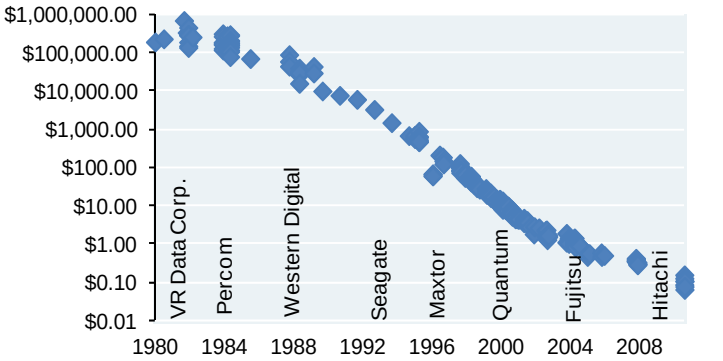
[2] We should expect more movies about machines and robots turning on humans, as in *Terminator 3: Rise of the Machines*, *I, Robot* and *Matrix Reloaded*. I never understood a single minute of any of the Matrix films.

A Bug's Life: Investment opportunities in Big Data

- ***Sensor nodes***, connected to the "internet of things". There are roughly 30 million sensor nodes connected to devices in transportation, automotive, retail and utility sectors, with the number of nodes growing at 30 percent per year.
- ***Healthcare databases***. Massive databases stored by healthcare companies, hospitals and device companies and government agencies, most of which until recently were not integrated or synthesized with each other.
- ***Non-healthcare corporate databases***. A wide range of Enterprise Resource Planning System and Customer Relationship Management databases. For example, there were 71 billion debit card and credit card purchase transactions in the US in 2011 populating ERP databases (*Nilson*), and 135 billion globally.

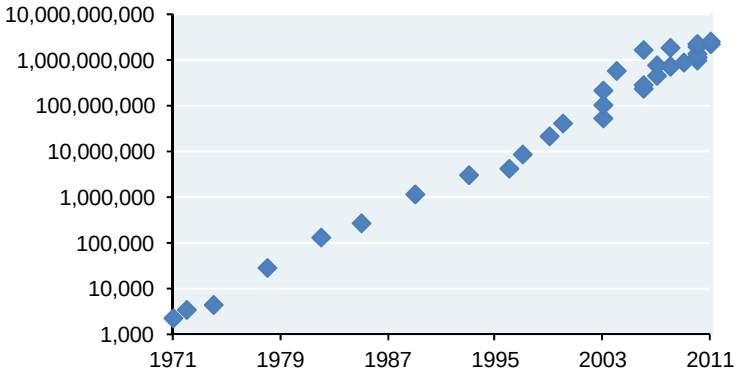
None of this would be possible without the growth in internet bandwidth, the collapse in the cost of storing a byte of data, and the increase in computing power. On bandwidth, the Broadband Forum reports over 600 million subscribers globally in Q3 2011. On the cost of storage, the first chart below shows one of the most spectacular declines in per unit costs you will find anywhere, measured as storage cost per gigabyte. The second chart shows the increase in computing power, measured as the number of transistors per processor. **Both storage cost and processing power charts are shown in log scale, an indication of the seismic shift that has taken place (they cannot be shown linearly since the gains have been so dramatic).**

Cost of a gigabyte of storage
USD, with indicative manufacturers



Source: Compiled by independent researchers, using individual manufacturers' suggested retail prices.

Processing power
Transistor count per processor



Source: Intel, AMD, Sony, IBM, Sun Microsystems.

A real life example: it is now possible to store 32 GB of data on a chip for your digital camera measuring 11 x 15 millimeters, weighing half a gram and costing under $100. This is millions of times lighter and 30,000 times cheaper than an equivalent device from 30 years ago. So, with the growth in high-speed bandwidth connections, the collapse in the cost of storing data and the proliferation of devices that generate it, let's get to the most important part**: how do companies use Big Data, and what kind of new businesses will arise to meet their needs?** Here is a brief review of some existing and emerging technologies which leverage Big Data:

Big Data products and services

Government • Eliminate income tax, sales tax or Medicare fraud and reduce unnecessary payments to vendors by reviewing Earned Income Credit, Dependent Child Care Credit, Itemized Deductions and withholding taxes • Defense and Security agencies analyzing data collected through satellites, signal intercepts and public sources • Municipalities using mobile phones for toll collection instead of separate transponders; and for identifying where potholes are (using smartphone accelerometers)	*Healthcare* • Digitization of medical records, funded with $20 billion by American Recovery and Reinvestment Act of 2009. There are 60,000 data elements per medical record as per Harvard Med School's CIO. From 2009 to 2011, the % of US hospitals that adopted electronic health records rose from 16% to 35% • Hospitals using "clinical decision support systems" to check for physician prescription data entry errors, reducing adverse reactions and related liability costs • Healthcare providers using remote monitoring devices (both data and video) to see if patients are following prescribed behaviors, and to prescribe treatments
Pharmaceutical • Analysis of clinical success and cost effectiveness of new drugs and treatments; and analysis of existing drugs, at times with the goal of pulling them off the market sooner if they're not working as planned. Vioxx is often cited as an example of where Big Data analysis might have resulted in faster recognition of its adverse effects	*Automotive* • Integration of traffic conditions, accident status, maintenance needs and service history. Notable examples: GM and BMW • Companies analyzing traffic patterns based on the aggregation of mobile phone data inside cars (using each mobile phone's cellular signal and the location of nearby cell towers), offering potential fuel and CO2 savings to users

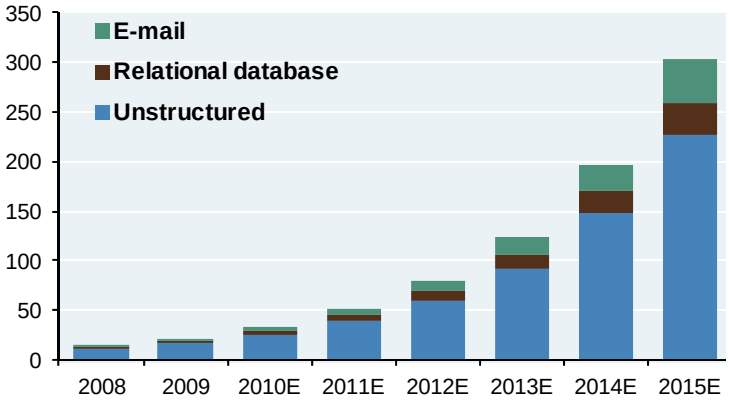
A Bug's Life: Investment opportunities in Big Data

Retailing • Radio frequency identification tags tracking movement of shopping carts and other in-store shopping behavior based on where smartphones congregate • Real-time customized recommendations to customers; advertising companies offering premium "geo-targeted" ads to retailers based on where potential customers are located. Other applications synthesize social media content with campaign spend to test its effectiveness • Retailers in particular need to benefit from Big Data, since other aspects of the internet are hurting them, such as real-time price discovery. Applications like RedLaser allow customers to scan barcodes with smartphones to get competitive pricing data. Web-based and web-influenced purchases will soon represent more than 50% of total US sales (*McKinsey, Forrester*). Another sign of pressure: in 1999, retailers earned more than half of all operating profit on goods sold to consumers, with the rest going to consumer goods products and packaging companies. Today, retailers only retain 30%	*Financials* • Brokerage firms operating in environments where information asymmetry is the key to success. This asymmetry is often associated with the speed of executing orders and processing information • Insurance companies using satellite imagery to assess residential or commercial real estate property risks to establish the right pricing • In 1990, checks and cash represented 84% of all purchase transactions. By 2015, debt and credit card purchases are expected to be 67% of all purchase transactions, adding to the flood of data (*Nilson*)
	Telecommunications • Use of social media to target customers-at-risk with retention plans, decisions based on whether their "friends" (based on public data) have already terminated a similar service
Aerospace, defense and semiconductors • Complex modeling applications to reduce production costs, particularly when final products are built from thousands of individual parts sourced from hundreds of individual suppliers	*Other* • Retailers, utilities and other service companies using route optimization applications to reduce mistakes and energy costs on customer deliveries; and to produce tire pressure alerts (reducing the risk of accident)

There's a technical aspect to this: how Big Data comes to life. There's little benefit to listing here all the tools for storing and analyzing massive data; most of us will never have heard of them (Hadoop, anyone?[3]). The important thing to understand is that a lot of data companies are now aggregating cannot easily be stored in structured, relational form**. Most is "unstructured" (text messages, machine data, images, social media feeds and video), and requires sophisticated tools to store.** To analyze the data, some tools are basic while others rely on complex machine learning skills to interpret data and have computers think for themselves. According to Carnegie Mellon, demand for expertise in machine learning far exceeds the supply, an imbalance which may become more severe (*note to parents of good math students who waste too much time playing World of Warcraft: there's a career out there for them*). After figuring out patterns and trends, programs then need to inform their human operators of what they found. Visualizations like clustergrams, history flow charts, and spatial information flow diagrams are designed to help humans understand what's going on. MIT's Senseable City Lab and GE partnered on a visualization project, part of which is shown below: an analysis of 217 million medical records, with the goal of finding patterns of co-occurring diseases (e.g., Depression and Tobacco abuse) that might not be apparent.

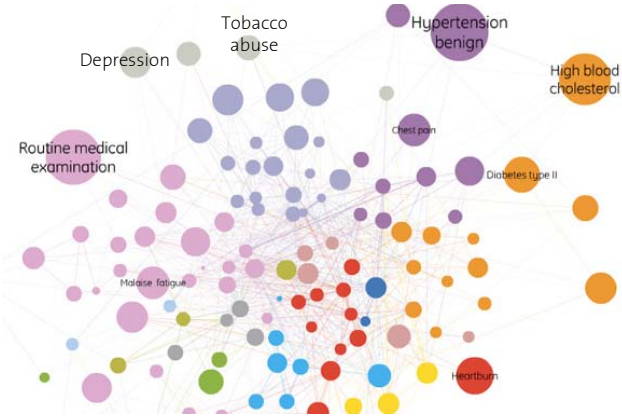
Unstructured data takes over

Worldwide total archived capacity, by content type, exabytes



Source: Enterprise Strategy Group.

An example of data visualization from GE and MIT



Source: MIT Senseable City Lab Health Infoscape project.

[3] Examples of unstructured database technologies include Splunk and NoSQL in addition to Hadoop.

A Bug's Life: Investment opportunities in Big Data

Before diving into specific examples, it's worth looking at how Big Data firms may be of interest to large technology companies. The table below (left) shows how some of the largest tech companies (Microsoft, Oracle, SAP, IBM, Symantec, EMC, CA Technologies, Adobe and Hewlett-Packard) have experienced declines in their valuation multiples over the last few years as earnings growth slowed. Given the need to find new sources of revenue and earnings growth, many of these very cash-rich companies have been acquiring Big Data and other technology companies (shown in the table on the right).

Decline in large cap tech valuations

	2006	2011	%change
Market cap ($bn)	$88	$84	-4%
Enterprise value/operating cash flow	13.5x	6.3x	-7.2x
Next 2 years of expected revenue growth	12%	6%	-6%
Next 2 years of operating cash flow growth	20%	12%	-7%
Number of > $100mm deals in last 5 years	5	7	40%
Value of > $100mm deals in last 5 years ($bn)	$7	$13	86%

Source: Bloomberg.

Recent acquisitions by large cap tech companies

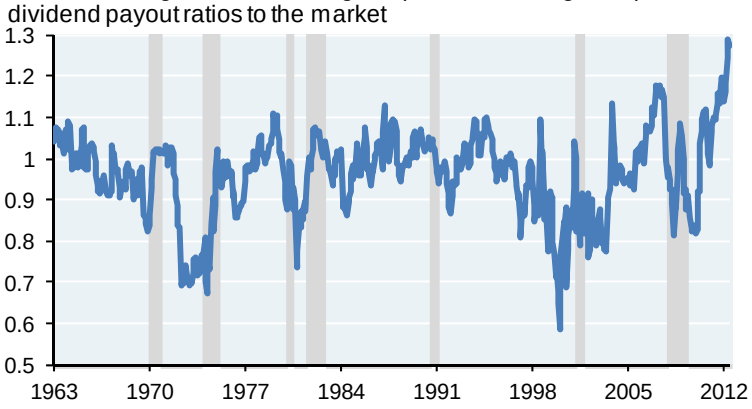
Acquirer/Target	Date	Price/Trailing 12m EBITDA	Price/ Revenue
HP/Autonomy	2011	23.9x	11.1x
Oracle/RightNow Tech	2011	63.7x	6.8x
IBM/Netezza	2010	85.1x	7.4x
SAP/Sybase	2010	13.0x	4.5x
Adobe/Omniture	2009	45.3x	4.6x

Source: Bloomberg.

There's another aspect of the current market environment that makes growth companies involved with Big Data potentially interesting: **investors have de-rated most growth stocks and left them for dead. Instead, investors are piling into income-generating stocks at the fastest pace seen in decades.** In prior notes, we highlighted how large cap technology multiples are flat to the broad market, in contrast to the last few decades when you had to pay a premium for growth. Now, in the second zero-rate environment created by the Fed over the last decade, there's another income frenzy going on. This has led to a rush into income-producing stocks (e.g., ones that pay high dividends). As shown in the accompanying chart from Mike Goldstein at Empirical Research, the P/E ratio of the highest dividend payers is at a record valuation premium compared to the P/E of the broad market. In this kind of environment, emerging growth companies may trade cheaply compared to stocks paying periodic dividends, but which are not growing as fast.

High dividend stocks: High relative valuations

Relative trailing P/E ratios of large-cap stocks with highest quintile of dividend payout ratios to the market



Source: Corporation reports, NBER, Empirical Research Partners Analysis.

Understanding Big Data and investing in it is a specialized discipline. As in the oil & gas sector, which we address in an *Eye on the Market* in March 2012, industry knowledge and experience matters. We tend to invest in this area with specialized, dedicated managers, rather than relying on generalist funds to get involved on an ad-hoc basis. With that backdrop, here are some public and private companies that managers we know have been investing in, each with a Big Data component. We define Big Data here liberally, and include companies for whom large data sets and analysis are central to their operations.

Example 1: **Reducing the cost of the diaspora of out-of-network doctor-patient relationships**

Most people covered by large US insurance companies use doctors that exist within their contracted network of providers. Our contacts suggest that this number is 80%-90%, measured as a percentage of all filed claims. The remainder represents decisions by patients to see out-of-network doctors. While out-of-network medical services seem pretty straightforward, like most things in the US healthcare system, they aren't. **Most doctors providing out-of-network services have to deal with patient credit risk[4], significant payment delays and administrative inefficiencies**. As for insurance companies, they typically cover a fraction of these out-of-network costs, and normally cap their exposure by using observed costs for similar procedures as the basis for what they will cover. Even so, they are **keen to reduce their payments to out-of-network providers used by patients they insure**, and also to reduce the resource-intensive costs associated with processing out-of-network claims.

[4] In some cases, insurance companies pay doctors directly, such that the provider's credit risk is the patient's portion only. However, in other cases, insurers pay the patient and providers bill patients for the entire amount, increasing the amount of the provider's credit risk.

A Bug's Life: Investment opportunities in Big Data

That's where a Big Data solution can help out, with an intermediary providing services of value to both doctors and insurance companies. As you might imagine, there's a lot of data involved here; Americans file over 2 billion claims per year for medical reasons[5]. Here's how it works, with the intermediary defined as HCDC (health care data company):

- HCDC recruits doctors to join its "out-of-network" network. After doing so, doctors agree to accept a discount to their "rate card" (standard fee-for-service rates), perhaps on the order of 20%-25%. For comparison, in-network providers are expected to provide discounts of around 40%. In exchange, doctors benefit from: (a) greater patient flow resulting from expanded patient access; (b) expedited payments; (c) reduced credit risk, since HCDC requires insurers to pay doctors directly, limiting credit risk to the patient portion only[6]; and (d) help with the data blizzard involved. **Cutting through the jargon, the doctor agrees to become part of a network that requires lower discounts in exchange for help dealing with a multitude of insurance companies, and lower patient credit risk.**
- HCDC's customers are commercial insurers, national and regional plans, self-insured employers, federal and state agencies and other entities which provide health insurance. Insurers agree to make payments to HCDC for a "match": that's when they can scan their universe of doctors to see if a patient's out-of-network provider is in there. If so, the insurer benefits from a discount to what that provider would normally have charged, and faster and more accurate transaction processing. HCDC receives a percentage of the insurer's savings as a fee for service. HCDC also earns revenues from smaller insurers who pay a fixed fee per covered employee per month.

The Big Data component: HCDC processed over 100 million claims in 2011, representing roughly $70 billion in gross claim charges. Their current network includes 5,200 hospitals, 125,000 providers of medical treatments and diagnostics (MRI clinics, blood work labs, etc.), and 740,000 healthcare professionals on one side, and 1,400+ insurance companies on the other. While the industry uses some standard codes and traditional relational databases, there are many free form entries which vary from doctor to doctor, or across insurers. HCDC improves its bottom line as it figures out better ways of understanding semi-structured data, which results in more matched claims.

Where might the company's future growth come from? First, healthcare expenditures are expected to grow at an average rate of 6.7% per year from 2015 to 2019. Because HCDC's revenue comes from a percentage of savings earned for insurers, they benefit from a rising healthcare cost environment [*see box below*]. A second source of revenue growth would come from an increased "match rate", which relies on the company's analytics and doctor network. Currently, HCDC matches 35%-40% of the claims it receives from insurers; the unmatched remainder are reverted at no cost to the insurer. Long-term potential opportunities involve penetration of related markets (workers' comp, no-fault auto medical claims, managed Medicaid), and the sale of aggregated data and related analytics to hospitals, physicians and pharmaceutical companies.

ObamaCare and its impact on medical costs. I don't want to get dragged into a firefight on this, but most sources I trust believe that the recent healthcare bill will not reduce costs, either in the short run or the long run. The fundamental purpose of the bill was to expand coverage to the uninsured, and the bill delivers on that promise. The decision to expand coverage to the uninsured before figuring out how to slow the trajectory of medical costs for the 85% of the population that *does* have health insurance will be left for future generations to assess, and pay for. The best quote I have seen comes from Alan Sager at the Boston University School of Public Health: "*The job of figuring how to cover uninsured people used up all the political oxygen that was available. They didn't have the energy for costs*".

An example: The Independent Medicare Advisory Board established by the bill may propose changes in reimbursement for physicians and hospitals, but its proposals may not ration health care, raise costs to Medicare beneficiaries, restrict benefits or modify Medicare eligibility criteria. For more information, see *Orentlicher* as referenced in sources.

[5] **Practicing medicine may get even more complicated**. The World Health Organization maintains a classification of diseases which is the most widely used system in the world. Most providers use ICD-9 which has 17,000 codes, and which is required for Medicare and Medicaid. Part of ObamaCare mandates a switch to ICD-10 which lists over 150,000 codes. Codes get as granular as non-venomous arthropod bites, alligator attacks, paper cuts, contact with swords, injuries from volcanic eruptions, and being accidentally shut in a refrigerator. This could serve as additional impetus for doctors to search for intermediaries to help them manage the data blizzard from out-of-network claims.

[6] Physicians and hospitals typically collect only about 50% of the post-insurance balance due from insured patients, and only 10% to 20% of the balance from self-pay patients. Across the health care sector, this results in almost $60 billion in bad debt each year.

A Bug's Life: Investment opportunities in Big Data

***Example 2:* Bringing the Wal-Mart data management experience to small and medium sized businesses**

In the table on page 1, you will find Wal-mart listed under petabyte examples, since that's how large its internal database is. Wal-Mart does not aggregate this data out of some obsession with large data sets: more likely, they see it as an indispensable tool to manage and monitor inventories, supplier relationships and customer buying preferences. Until a few years ago, it was costly to deliver such tools to small and medium sized businesses. As a result, most of them just accepted customer payments via cash or credit, after which all the transaction data vanished. Such businesses would manually process their books and inventory, often using the same tools they might have relied on 20 years ago.

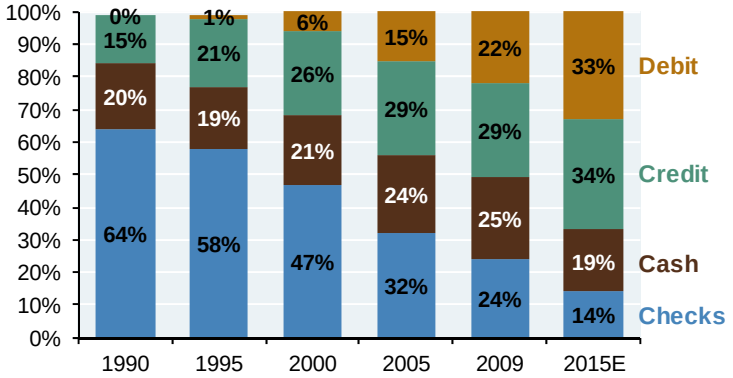
With the decline in processing power and data storage costs, an industry has emerged to provide payment processing solutions to small and medium sized businesses (SMBs). The pitch is straightforward: instead of spending $300 on a "dumb" credit card terminal, SMBs are offered integrated point-of-sale systems that help them manage their businesses. The merchant benefits from reduced data entry errors; improved reporting, inventory and cash flow management; and the ability to build out customer loyalty programs, web-based outreach programs and targeted advertising (using an SMB equivalent to Amazon's "recommendation engine"). The cost to the merchant is in thousands rather than hundreds; it's meant less for corner pizza parlors and convenience stores, and more for businesses with at least a few hundred thousand dollars in annual card volume.

One of the managers we know invested in a company that processes payments from hundreds of different types of smart terminals. The terminals sound basic pieces of hardware (a PC or a tablet), and they are. What makes it an interesting Big Data opportunity is the software and analytics which the hardware employs. The systems must seamlessly handle all kinds of payment types, such as credit cards, debit cards and gift cards, and importantly, interface with the merchant's own inventory and sales information in a secure and compliant fashion. The company processes around 1 billion transactions every year, each generating revenue (this revenue is in addition to whatever fees are paid to credit card networks and credit card companies). To accomplish all of this, the company has built a partnership with hundreds of software developers and thousands of independent local distributors which sell these integrated systems to SMBs.

One of the important drivers behind this kind of investment is the ongoing shift from cash and checks towards credit and debit payments (see chart, right), which makes these kind of integrated systems more valuable to merchants. Most of the 6 million SMBs in the US have some kind of credit card terminal, but the majority are the "dumb" terminals described above. The company described above is not large, and employs less than 700 people. The risk with a smaller player is that large payment processors have a lot more merchant relationships and a lot more scale, and may also seek to make inroads with their own smart terminal systems. So far, the company has been successful in growing its smart terminal merchant base at more than 20% per year, and is on the verge of becoming a top-ten payment processor by volume.

How consumers pay for what they buy

Percent



Source: The Nilson Report.

Example 3: whom to call if you have a terabyte of data to analyze

The first two examples dealt with industry solutions in healthcare and retailing. Other Big Data companies offer storage and analytics to companies across sectors. One of our technology managers has invested in the past in a company which offers integrated data storage, analytical and consulting services to multiple companies with Big Data needs. Here are a few of them:

- A European vehicle manufacturer needed help resolving discrepancies between warranty claims and computer-generated vehicle errors, which were housed in separate systems. In other words, do they have faulty computer chips or faulty mechanics? The company also wanted to be able to complete a comprehensive report of diagnostic failure codes by model and year in 15 minutes, down from its prior timeframe of 2 weeks
- A national casino chain wanted to integrate customer, gaming and hotel data in order to generate on the spot targeted promotions to highest-value guests (i.e., the ones who lose all their money[7]), and compare performance across properties
- A large railway operator wanted to optimize use of its equipment by analyzing whether it's offline, online, at the station, or in storage. The operator launched a program with guaranteed arrival times, which it is now meeting 98% of the time

[7] A 2011 Stanford Business School paper analyzed 2 years of detailed gambling records, and concluded that 8% of gamblers are addicted, while the remaining 92% gamble for entertainment purposes. Around 2% of gamblers are responsible for 25% of wins and losses.

A Bug's Life: Investment opportunities in Big Data

- One of China's largest commercial banks used the company's help to build an integrated data warehouse of **accurate** customer information from multiple sources (*accurate* being the operative word these days as it relates to Chinese data integrity). The data resulted in a marketing campaign which achieved a record 56% response rate on a new banking product.

The company has the ultimate in blue chip client lists, among them the world's largest financial, telecommunication, travel, transportation and retail companies. **What explains their success so far?** In addition to being able to handle data sets measured in terabytes, the company uses technology which is well-suited to clients with rapidly growing data storage needs, and with the need to perform high-level analytics (see the first box below if you are REALLY interested in the details). Another difference lay in its choice of analytical tools, as it relies on a higher-level approach better-suited to management-level questions, often involving unstructured data (for the truly intrepid, see the second box).

WARNING: do not read the following 2 boxes unless you are the kind of person who has taken apart a computer and then reassembled it

The storage and architecture wars. As companies grow, they are like parents with growing children: every once in a while, you need to buy new clothes. The difference is that children grow in somewhat linear fashion, while today's large, global enterprises can see their data storage and analysis needs growing exponentially, particularly as they try to mine data they never looked at before. The incumbent Big Data technology architecture is referred to as Symmetric Multi-Processing (SMP). In an SMP environment, each central processor core can work with any section of memory or disk, and all memory and disk is available to each core. The processor connects to the memory and disk by what is known as a memory bus. The challenge is that the memory bus can get overloaded with massive data sets; it becomes a bottleneck, creating the equivalent of a data traffic jam. As a result, buying bigger, more expensive servers does not result in a linear improvement in processing speed or analytical capabilities.

The newer alternative: Massively Parallel Processing (MPP). This approach uses multiple nodes (servers), each of which has its own memory and disk, allowing the workload to be shared. This way, the company can offer its clients increased storage and analytical firepower on a more cost-effective basis, and one where increased costs are better aligned with improved computing power. The idea of parallel processing has been around for decades, but the use of MPP for Big Data solutions is a newer phenomenon. The company described above connects multiple commoditized servers to meet client needs, rather than having to upgrade them to more and more expensive servers as their needs grow.

What Big Data processing approach do you prefer? For day to day Big Data queries, most companies use On Line Transaction Processing (OLTP). The data is updated frequently, with a huge premium put on data integrity. Its success is often measured in the number of transactions that can be processed per second. However, OLTP is not ideal for management-level queries about trends related to margins, customer behavior, supplier costs, etc. The alternative is On Line Analytical Processing (OLAP), which often relies on OLTP databases for basic information. Problem-solving on a multi-dimensional basis, rather than speed of execution, is the objective. Here's an example: suppose a hurricane is coming. Based on prior experiences, what products should a company have plenty of in reserve to satisfy its customer needs? This is an OLAP question, rather than an OLTP one.

The company's solutions are highly "sticky" and seldom removed, as building out an integrated data warehouse is extremely costly. Customers typically set up new data warehouses as new projects are launched. As customers' data grows and they run out of capacity, they tend to purchase additional products every 12-18 months. The company's customer base is broadly diversified with no single customer accounting for more than 10% of revenues over the last few years. Customers pay for hardware, a software license, and also pay a percent of the license fee each year for maintenance and support. The company's 2011 revenue mix was 23% maintenance, 29% consulting and 48% product (hardware and software).

Final comments

We have tried to shed a little light on the companies that store, manage and analyze Big Data, and what the investment opportunities might look like as the digital age marches on. There's a *lot* more to the broader Big Data universe, such as cloud computing, growth in mobile devices, the growth in social media and other trends which revolutionize the way data is created, stored and analyzed. If the world ever gets past its macroeconomic sand traps, perhaps we can focus on some of them. Big Data is interesting, since at a time when a lot of things appear to be slowing down, some of these companies are generating revenue growth of 20%-40% and cash flow margins of 25%-50%. Of the growth-oriented investments we make, these are among the ones we prefer. It goes without saying, but the **pitfalls and periodic over-exuberance in the technology sector** should always be present in anyone's mind. In our view, this kind of thing is worth including in a portfolio, but in manageable "byte" sizes.

Michael Cembalest
J.P. Morgan Asset Management

A Bug's Life: Investment opportunities in Big Data

Sources

- *"Big data: The next frontier for innovation, competition, and productivity"*, McKinsey Global Institute, June 2011
- *"Big Data Primer: Laying the Foundation for Broader Analytics"*, J.P. Morgan North America Equity Research, June 2012
- "*Masters of Big Data: Concentration of Power Over Digital Information*", Alessandro Mantelero, Polytechnic University of Turin - Department of Production Systems and Business Economics, February 2012
- "*Data Science and Prediction*", Vasant Dhar, Stern School of Business / Center for Digital Economy Research, March 2012
- Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011–2016
- "*Shifts in U.S. Consumer Payment Systems*", The Nilson Report, Issue # 962, December 2010
- *"Overhauling the US health care payment system"*, The McKinsey Quarterly, June 2007
- "*Cost containment and the Patient Protection and Affordable Care Act*", David Orentlicher, Indiana University, as published in the Florida International University Law Review, 2011
- "*Big Data Meets Big Data Analytics*", SAS White Paper
- "*Implementing a Microsoft SQL Server Parallel Data Warehouse Using the Kimball Approach*", SQL Server Technical Article, Microsoft, June 2011 (for descriptions of SMP vs MPP architecture)
- Source for cost of hard drive storage over time: http://ns1758.ca/winch/winchest.html