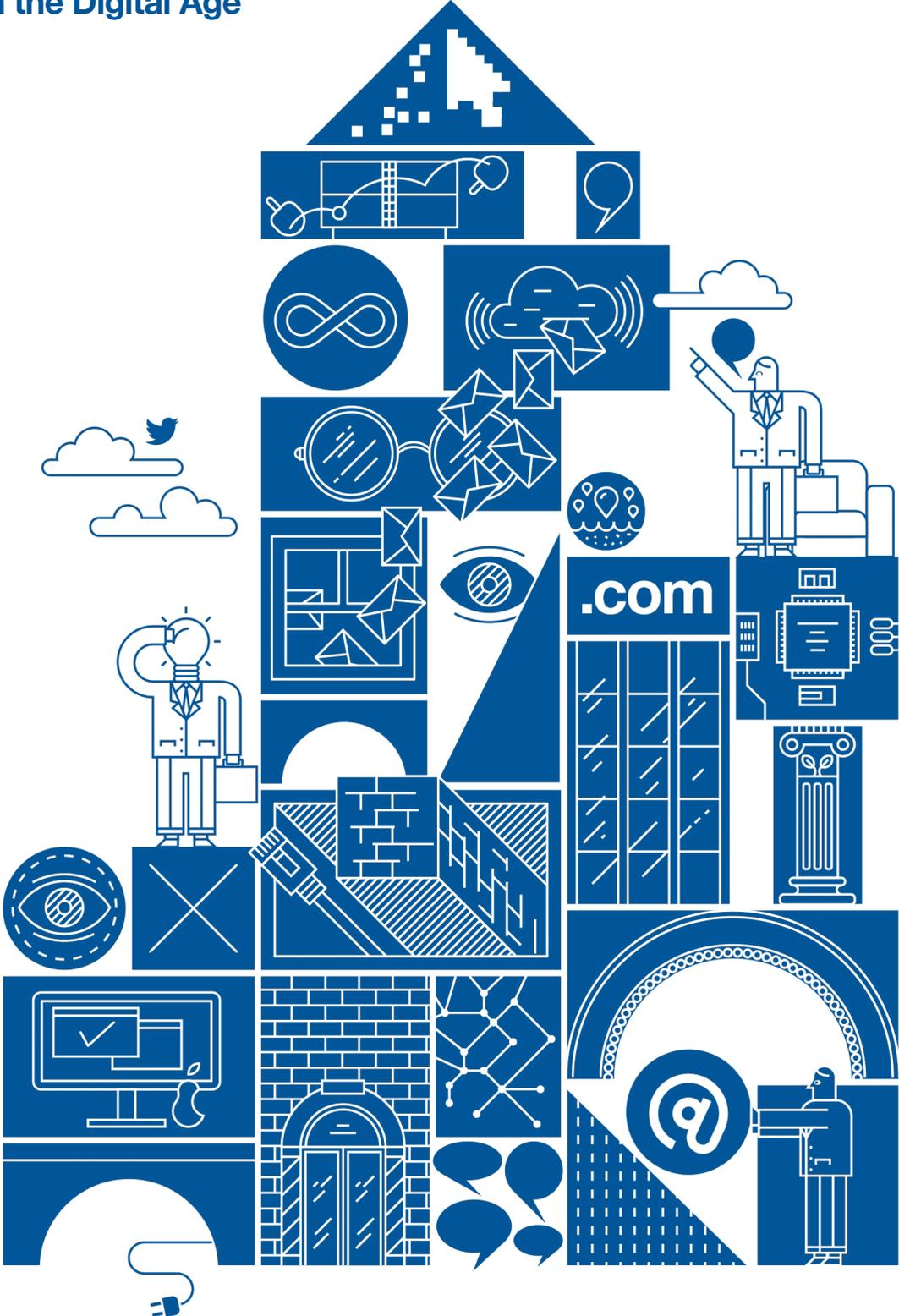


Reinventing the Company

in the Digital Age



Reinventing the Company

in the Digital Age

From Deconstruction to Big Data: How Technology is Reshaping the Corporation

Philip Evans

Evans affirms that we are undergoing a re-acceleration of technological change despite the global recession and that something sudden and dramatic is happening. One important aspect of this is how Big Data is reshaping business, and transforming internal organization and industry architecture. He goes on to explain that two information technology drivers are reshaping internal organization: business strategy and the structures of industries. The first is deconstruction of value chains: the breakup of vertically-integrated businesses, as standards and interoperability replace managed interfaces. And the second is polarization of the economies of mass, meaning that in some activities, economies of scale and experience are evaporating, while in others they are intensifying. He doesn't consider Big Data as an isolated or unique phenomenon, but rather as an example of a wider and deeper set of trends reshaping the business world. In his article he describes how the broad logic of deconstruction and polarization of scale can be applied to the specific case of Big Data and the corporation, and finds that these apparently contradictory strategies are mutually complementary.

Philip Evans

Boston Consulting Group



Philip Evans is a Senior Partner with the Boston Consulting Group and BCG Fellow. He founded BCG's Media and Internet practices. He was educated at Cambridge University, was a Harkness Fellow at Harvard, and obtained an MBA from the Harvard Business School. He is a Director of the Oxford Internet Institute. He is author of many publications, including "Strategy and the New Economics of Information" which won a McKinsey Prize, awarded annually for the best contributions to the *Harvard Business Review*. His book *Blown to Bits: How the New Economics of Information Transforms Strategy* (1999) was a global best-seller.

Key Features for the Company of the Future:

Organize the Business Along Its Potential Fault Lines

The most important enablers of "disruption" are standards and interoperability, which break prevailing value chains. The correct response is to organize around that contingency, loosely coupling successive value-added steps and defining organizational units by their distinct competitive economics. Increasingly this means substituting a more functional organizational structure. This gives managers a clear line-of-sight on the threats they face, and gives senior managers the option to redraw business boundaries and make incursions into the business models of others.

Devolve Key Activities to Autonomous Communities

Within such a loosely-coupled, functional structure, some activities emerge as better done by communities of autonomous individuals. These can be users, independent experts, or enthusiasts, most frequently one's own customers. They can also be communities within the corporation itself: engineers and front-line staff coming together autonomously to swarm over glitches and innovate on features and interfaces. This requires a different mode of management: curatorial rather than hierarchical, enabling rather than directing.

Build Shared Infrastructure

Both within the corporation and across its boundaries many functions are scaling beyond traditional business unit boundaries, possibly beyond corporate boundaries. In particular data and the distributed commodity computing performed on large data sets, become infrastructure: a resource shared across the company. Private and public cloud computing, and industry-wide data sets built as commons, will be essential to new services in many industries. With personal data, where data rights are crucial, this may require that data repositories be mandated as trustees to protect integrity and privacy.

From Deconstruction to Big Data: How Technology is Reshaping the Corporation

Businesspeople everywhere grasp that something sudden and dramatic is happening. Here are five salient observations.

- The number of transistors on an integrated circuit still doubles every two years. Storage density doubles every 13 months. The amount of data transmittable through an optical fiber doubles every nine months.
- Broadband internet access in the G-20 is growing from 800 million (of which 50% mobile) in 2010, to 2.7 billion (of which 80% mobile) in 2015.¹ The number of cellphones in the world is now equal to the number of people. 1-2 billion more people in the world have a cellphone than have a bank account—or a toilet.² Smartphone sales reached one billion units in 2013 (up 66% over 2012). Smartphones are the fastest-adopted technology ever.
- Facebook has 1.3 billion active users. 64% visit the site daily (spending an average of 20 minutes). 4.5 billion “likes” are posted daily.³ Half

a trillion photographs are uploaded to the web each year, and one hundred hours of video to YouTube every minute.

- The number of IP-enabled sensors will exceed 50 billion by 2020.⁴ RFID tags now cost as little as 5 cents. Estimates vary, but the range of projections is for the total number of sensors in the world to reach one to ten *trillion* between 2017 and 2025.
- 90% of the world’s stock of data was generated in the past two years.⁵ 99% of that is now digitized, and over half IP-enabled, meaning that technically it can be uploaded and shared over the internet. Half of the world’s knowledge is potentially a single document.

Most of this is really sudden: a re-acceleration of technological change that seems to have occurred in the last decade, after the lull of the dot-com bust, and despite the global recession. It is deeply disorienting: people speak of “disruptive technologies” meaning change which incumbents—by definition—cannot deal with. Managers in established companies crave something more specific than the proposition that they are destined to be “disrupted” by some kids in Silicon Valley. But with the current pace of change it would be a rash person who claimed to be able to forecast the fate of specific businesses or corporations: Apple, for example, has been declared “dead” by commentators in the press 64 times since April, 1995.⁶ At the time of writing it is the world’s most valuable corporation.

90% of the world’s stock of data was generated in the past two years. 99% of that is now digitized, and over half IP-enabled

To cope with this degree of fluidity and uncertainty, the strategist needs to return to first principles. We cannot assume that traditional bases of competitive advantage will last. We cannot presume that hard-earned “excellence,” built within the current business model, is the right skill-set for the future. We do not know who our future competitors will be. Indeed the boundaries of the business and the industry cannot be taken for granted. We need to step back and rethink the connection between technology and business strategy.

I believe that the general principle is as follows. Two large phenomena, both driven by information technology, are reshaping internal organization, business strategy and the structures of industries. The first is *deconstruction* of value chains: the break-up of vertically-integrated businesses, as standards and interoperability replace managed interfaces. And the second is *polarization of*

the economies of mass, meaning that in some activities, economies of scale and experience are evaporating, while in others they are intensifying. “Negative” polarization, where economies of scale and experience have weakened, leads to the fragmentation of activities, often to the limiting case of individuals in communities replacing corporations as the principal actors. “Positive” polarization, where they have strengthened, leads to the concentration of activities, often to the limiting cases of utilities, co-ops or monopoly. The combined consequence of these trends is to substitute “horizontal” organization for “vertical,” both within the corporation and across industries. The transposition of the industrial matrix.

Deconstruction of value chains and polarization of the economies of mass are reshaping internal organization and the structures of industries

This does not render the traditional corporation obsolete, but it does often mean that corporations need to redefine their role and reshape their business definitions. They need to establish collaborative relationships with communities, especially user communities, where individuals or small proprietorships are more flexible, better-informed about end-use, or can innovate more cheaply. Conversely they need to establish collaborative relations with other institutions, perhaps competitors, to achieve economies of scale and experience that would otherwise be inaccessible. On both sides, strategy becomes a matter of collaboration as well as competition.

Internally corporations need to do much the same thing. Innovation and small-scale experimentation are best done in loose groups where individuals and small teams enjoy a high measure of autonomy. Conversely scale- and experience-sensitive functions need to be centralized across businesses, driving the overall organization to a more functional structure. The internal architecture of the corporation becomes a set of platforms, each supporting activities at smaller scale and with faster cycle times. One platform can be stacked on top of another. And the architecture of an “industry” can be exactly the same, some companies serving as platforms for others, some serving as platforms for end-user communities. The pattern is fractal.

These trends are quite general, and account for numerous industry disruptions. But they apply in particular to Big Data. “Big Data” means much more than vastly larger datasets and exotic software. It requires treating data as infrastructure: centralized, secure, massively-scaled, built as a general resource not for any specific end-use. It also requires treating the processes of inference as “super-structure”: iterative, tactical, granular, modular, decentralized. Put the two together internally and you are replacing product- or market-based organization with a functional one. Put the two together externally and you have a fundamental challenge—a disruption—to many traditional business models.

Thus, Big Data is not an isolated or unique phenomenon: it is an exemplar of a wider and deeper set of trends reshaping the business world. Achieving the potential of Big Data is a challenge not only to process and capabilities, but also to organization and strategy. It is an issue for the CEO.

In this chapter I plan to survey the broad logic of deconstruction and polarization of scale, and then apply it to the specific case of Big Data. I hope that by stepping back in this fashion we can see its longer-term strategic and organizational significance.

Deconstruction

Activities can be vertically integrated for two possible reasons: the *technical* need to coordinate a complex or ambiguous interface, and/or the *moral* need to align the interests of the two parties, without contracts and lawyers. Technology weakens both rationales: as economists would put it, technology lowers transaction costs.

The fundamental technical drivers, of course, are the “Big Exponentials”: the falling costs of computing, storage, and communication. The first-order consequence is that both parties to a transaction have access to far more (and more timely) information about each other, and about alternatives. Search, comparison, benchmarking, qualification, price discovery, negotiation, and auditing all become orders-of-magnitude cheaper and more comprehensive. In the context of this explosion in reach, the logic for standards becomes compelling: simplifying interfaces, setting mutual expectations, promoting interoperability, and nurturing the network effect. By commoditizing interfaces, standards reduce, often eliminate, the need for technical coordination.

The moral argument is a bit less obvious. Information asymmetries inhibit transactions (“what does the seller know about this used car that I don’t know?”) Technology generally increases the information symmetry between transactors. So technology can reduce the economic inefficiencies stemming from rationally defensive behavior by the less-informed party. When the repair history of a car can be read from a data socket under the dashboard, buyer and seller can close a deal with much greater ease.

Further, electronic technologies can put transactors in front of a virtual audience. The rating systems curated by Amazon, Etsy, and Yelp give each product or seller a cumulating “reputation” which is a surety for trust. Amazon encourage customers to rate not just the products, but the raters, awarding stars and badges to their most frequent and consistently constructive contributors. The more broadly visible and persistent the reputation, the more an individual can be trusted to act to preserve it; the higher the trust, the lower the need to negotiate, monitor, see for oneself, write and enforce a contract. Reciprocity is

social capital established between two parties: it “hard wires” trust because it requires the investment of multiple transactions between those parties for the mutual trust to be established. Reputation, in contrast, is portable within a community: trust earned in one context can be relied on in another. Reputation

Electronic technologies can put transactors in front of a virtual audience. The rating systems curated by Amazon, Etsy, and Yelp give each product or seller a cumulating “reputation” which is a surety for trust

“soft wires” trust. Technology enables a wholesale switch from reciprocity to reputation, embeds reputation in data, and allows reputation to scale beyond the traditional limits of geography or institution.

Transaction costs serve as a sort of “set-up” cost for a transaction. So, lower transaction costs reduce the threshold transaction size, making it possible to execute smaller, more granular transactions (eBay started

as a mart for Pez dispensers). And this feeds on itself: the smaller the transaction, the less the gain from opportunistic acts relative to the reputational risk of being caught taking advantage of a counterparty. People and companies have, therefore, stronger reasons to avoid opportunistic behavior; *other* people have, therefore, stronger reasons for trusting them. Transactions throw off data, data sustains trust, trust enables transactions: a virtuous circle.

Visibility lowers transaction costs by another mechanism increasingly relevant to Big Data: it creates a “negative cost” to transactions, derived from the value of the information generated as a byproduct: the “data exhaust.” As long as the parties that are the subject of the data are indifferent to its ancillary uses (an important caveat!), this beneficial offset lowers the *net* cost of transacting. When this positive value is sufficiently high, it can warrant providing the underlying service for free, just to capture the transactional data. This, of course, is the model of many internet services, notably search and social networking. Freeness in turn eliminates another tranche of transaction costs that would otherwise be necessary to maintain accounts, invoice, and collect. (Half the cost of the phone system, for example, is *billing*.) Whether the transactors are (or should be) indifferent is a different question. Just as transparency can *create* trust, so transparency can *require* trust: trust in the entity collecting and using the data.

Exactly how this logic plays out varies, of course, from one domain to another. But the themes are as predictable and recombinant as the ingredients in a Chinese menu: standards, interoperability, information symmetry, reputation-based trust, “free”; all in the context of cheap global connectivity. The pervasiveness of the Big Exponentials, and their relentless downward pressure on transaction costs, result in the universal weakening, and frequent melting, of the informational glue that holds value chains together. This is *deconstruction*.

Polarization of Economies of Mass

Businesses in a traditionally structured industry compete on similar, vertically-integrated value chains comprising a bundle of heterogeneous, roughly sequential activities: sourcing, machining, assembly, distribution, advertising, etc. Advantage in one element might well be offset by disadvantage in another. Many activities exhibit increasing returns to scale and/or experience (which I lump together as “mass”), but many do not. There might even be activities with negative returns to mass: where bigger simply means loss of flexibility and more overhead. This is why, *averaged across all the components of the value chain*, we have typically seen only gently increasing returns for a business as a whole. Therefore, in a maturing industry, multiple competitors could survive, their profitability positively (but not overwhelmingly) correlated with market share.

But deconstruction, by ungluing different value-chain steps and allowing them to evolve independently, undermines the “averaged” pattern of gently positive returns to mass. Instead, each step evolves according to its own laws.

Where economies of mass are *negative* the activity will fragment, perhaps into a population of small proprietorships, such as the developer and producer communities that flourish on such platforms as iOS, Alibaba, and Valve. In the limiting case, autonomous individuals come together in communities for the purpose of “peer production” of information goods. Users of the good or service are often those most motivated and best positioned to make improvements for their own purposes, and if the contribution in question is information, sharing their improvements is costless to the sharer. Contributions can be in such small increments that non-financial motivations—whether fun, altruism, reputation, or applause—can suffice. Maybe it is merely because people are willing to donate their labor, maybe because tasks can now be cost-effectively broken down into smaller pieces, maybe because hierarchical management in some circumstances is merely overhead, maybe because there is some ineffable and emergent phenomenon of collective intelligence: *it works*. Hence Wikipedia, hence Linux, hence the body of reader reviews on Amazon: coherent intellectual edifices built from thousands of autonomous and unpaid contributions.

What is new here is not the possibility of productive communities (they are, after all, a tribal mode of coordination that antedates both markets and hierarchical organization), but rather the new ability of communities to *scale*.

Deconstruction, by ungluing different value-chain steps and allowing them to evolve independently, undermines the “averaged” pattern of gently positive returns to mass

With scale comes complexity, emergent structure, and the gravitational pull of the network effect. For certain kinds of production, globally scaled communities not only get stuff done, but are economically advantaged over traditional corporate hierarchies and markets in doing so.

Where economies of mass are strongly *positive*, the reverse logic applies: the activity concentrates and may indeed become a monopoly. Sometimes the economies of scale were always present but locked inaccessibly within the value chains of competing corporations. Sometimes, as with fiber optic networks, genomic science, cloud computing—and of course Big Data—the scale economies have emerged in consequence of new technologies.

Deconstructing Data

So how does this logic affect “data”? The short answer is that digitization—which is largely complete—permits deconstruction, and we are now entering the era of polarization. Economies of mass—both scale and experience—are polarizing in favor of the very large: that is “Big Data.” But they are also polarizing in favor of the very small, as teams and individuals become the vehicles to extract “Big Insight.”

Data was the by-product of other activities. It was analog and short-lived: generated and consumed on the spot, or passed along value chains like (indeed *as*) *kanban* tickets on a Toyota assembly line. Most often it was then discarded, or if retained, filtered and formatted in rigid schemas such as accounting, for narrow and predetermined purposes.

Data, like all information, has a fixed cost of creation or collection, so even prior to digitization it was amenable to economies of scale through the amortization of that fixed cost. And the logic of statistical inference has *always* dictated that larger data sets yield superior insight, whether in the number of patterns or discriminations that can be inferred at a given level of confidence, or in the confidence with which a given conclusion can be drawn. But until recently these scale and experience economies have not predominated because of constraints in collection, storage, transmission, processing, and analytical technique. We worked with smaller datasets because we could not cost-effectively gather all the data, array it, and do the sums. Scale and experience economies inherent in data were locked inside processes, places, and value chains.

But digitization drove the cost of data replication to zero, communication drove the span of replication to the universe, and the cost of storage is falling by a factor of a thousand each decade. The “Internet of Things” is how we now gather data, ubiquitous mobility is one of many ways we both produce, transfer, and consume it, and the cloud is the architecture of storage and computation.

Economies of “mass” are extended: scale economies from exploiting the flows of data, and experience economies from exploiting the cumulating data stocks.

“Data wants to be Big.” Finally, technology makes that possible.

Consequently, minimum efficient scale for data and the facilities that house it is growing, first beyond the reach of individual business units within the corporation, and ultimately in many cases beyond the corporation itself. Hence cloud computing and remote data centers: first within the corporations, then outsourced to providers such as Amazon enjoying even greater economies of scale. As data hyperscales, it becomes rational to treat it as infrastructure: general in purpose, capital-intensive, supporting multiple activities. It becomes long-lived, as much a stock as a flow.

But the collection of data, in itself, is of very limited value. The valuable thing is the *insight* that can be derived from the data. “Big Insight” requires that the analytical process scale along with the Big Data that it uses. Since the complexity of analysis is often far more than proportional to the number of data points employed, our ability to do analysis on very large data sets is not guaranteed by the progress of the Big Exponentials. A Cray supercomputer running traditional analytical methods at staggering speed is not the solution to the problem of analyzing immense data sets: beyond a certain throughput the machine simply melts. Instead statisticians and computer scientists have developed two new strategies to enable the scaling of insight.

The first is *iteration*: instead of striving for a formal and complete solution to an analytical problem, they construct computationally simpler algorithms that guess at the answer with progressively increasing accuracy. Any estimate, indeed the truth-value of any data point, is merely interim, subject to emendation or correction as new data points are collected. In essence, inference becomes a “Bayesian” process of revising probability estimates as new information is incorporated. And inference becomes a process rather than an act: instead of solving the problem once, the solution is approximated and re-approximated continuously.

Statisticians and computer scientists have developed two new strategies to enable the scaling of insight. The first strategy is *iteration* and the second is *decomposition*

The second strategy is *decomposition*: solving a large problem by breaking it into many small pieces that can be computed in parallel. This is a rapidly developing branch of statistics: finding new ways to solve *in parallel* problems that traditionally have been solved *sequentially*. Such solutions can be calculated, not with a supercomputer, but with racks of cheap, low-performance

commodity servers. So data centers, with hundreds of thousands of such servers, become repositories not just of Big Data but also of computing Big Insight. Instead of the data going to the query, the query must go to the data.

Together, iteration and decomposition allow insight to scale. The “poster child” example of Big Insight is Google Search. The underlying problem is to calculate the “centrality” of each page of the World Wide Web, as defined by the number of other pages pointing to it, but weighting each pointing page by its own centrality score. Mathematically this is the calculation of something called “eigenvector centrality,” a trivial piece of linear algebra. The problem is that the number of arithmetical operations required to solve it is proportional to the *cube* of the size of the World Wide Web: with four and a half billion web pages, it cannot be done. Larry Page’s inspiration was to develop an

Together, iteration and decomposition allow insight to scale. The “poster child” example of Big Insight is Google Search

algorithm that approximated the solution to this problem well enough for practical purposes. That is PageRank. To implement the algorithm Google runs a crawler: software that searches the internet continuously for new web pages and links. The content of the web pages and their locations are continuously re-indexed and stored in literally millions of servers: each server might contain, for example, a list of the addresses and PageRanks of every web page that contains a given word. When you or I perform a Google search, the heavy work is done by an instance of a program called Map/Reduce, which decomposes our query into its constituent words, sends those queries to the relevant index servers and then reassembles the results to sort the pages most likely to satisfy our search. The Map/Reduce program does not need to know where a specific index resides: instead there is a “virtualization” layer of software, called Big Table, which stands between the Map/Reduce programs and the index servers. Big Table adds and backs up servers, reassigns data among servers, and works around machines that fail, all without the Map/Reduce software needing to know.

Three principles: data-as-infrastructure, iteration and decomposition. In Google Search they work together to solve problems unsolvable by conventional methods, and do so at global scale. And in a quarter of a second. This may seem alien and exotic, but it is merely a pure case of three principles that apply in every corporate environment.

Google Search has an important complementary consequence: it *removes* traditional economies of scale and experience from the process of searching. The searcher does not need to be a professional librarian and does not need to be located in a research institution. All the searcher needs is an internet connection and a browser. So what was a profession, or at least a serious time commitment,

becomes a casual activity available to all. Within Google's own architecture the same thing is true: at low cost, Google can add new algorithms such as Spellcheck and Google Translate that sit on top of Big Table and tap into precisely the same data and computational infrastructure. Small and self-directing teams of engineers can experiment with new products and services, relying on the index servers and Big Table to do all the scale-intensive heavy lifting.

Google expose this architecture to outsiders. They have published about seventy APIs (Application Programming Interfaces) to make Google resources freely available to anybody with a website and simple programming skills. That is how your local restaurant uses a widget from Google Maps to provide driving directions on its web page. In all, some 12,000 APIs have been published by various companies. There is a cottage industry that has produced some 6,000 so-called "mashups" by combining these APIs to create new, small-scale services. These services may be small businesses, they may be hobbies, they may be fads, but it does not matter: precisely because the required resource commitment is so small, the cost of experimentation and the cost of failure have plummeted. The Very Small flourishes on top of the Very Large.

This is how Big Data emerges, not just as a new set of techniques, but as a new architecture for businesses and for industries. Interoperable interfaces such as APIs and Big Table allow different functions to evolve in accordance with their separate economics: they "deconstruct" the traditional value chain of linear inference. Once these interfaces are in place, scale-intensive assets (most notably data and data centers) and scale-intensive activities (most notably large, decomposed computations) can be centralized and managed for efficiency, capacity utilization, security and reliability. Indeed the performance of scale-intensive analytics can (and increasingly, must) be co-located with the data in the data centers. But conversely, tinkering with algorithms, the combination and recombination of different information resources to meet specific needs, and experimental inquiry, are all drained of their scale-intensity: anybody can do it anywhere. The cost of trial-and-error, replication, and redundancy become negligible. The overall "ecosystem" thus exploits the symbiosis between these two kinds of activities: infrastructure managed for efficiency, and communities self-organizing for innovation, customization and adaptability. The classic trade-off between efficiency and innovation is radically finessed.

So communities, cottage industries, amateurs, self-organizing teams, hobbyists and moonlighters, flourishing on immense platforms provided by the likes of Google, can now compete against the professionals in traditional

The Very Small flourishes on top of the Very Large. This is how Big Data emerges, as a new architecture for businesses and industries

organizations. The typical corporation is thus challenged on two fronts: by swarms of individuals and small groups which can innovate, adapt, and experiment at lower cost, and simultaneously by organizations which have a scale and experience level beyond its grasp. The typical corporation may simultaneously be too big and too small.

Too Big: Tapping the Power of Communities

Companies can address the problem of being too big, slow, and cumbersome by exposing their data to the energies and imaginations of external communities. That is what Google do with their web APIs, and Amazon with their customer reviews. (And those companies are no slouches!) This is *risky*: intellectual property may be compromised, and privacy must be protected. Retailers like Amazon risk specific sales from publishing negative reviews, outweighed, they hope, by the greater trust and credibility of the store overall.

One way to tap the energy of communities is through contests. In 2006 Netflix launched a contest to improve its movie recommendation engine. They released an immense, anonymized data set of how some half million customers had rated some 20,000 movies. Netflix promised a grand prize of \$1 million to whoever could first improve on their in-house recommendation algorithm by 10%. Intermediate prizes were offered for the best algorithm to date, conditional on partial release of the solution to other contestants to stimulate further innovation. Netflix thus cleverly set up a rich environment for both competition and collaboration. Over three years teams competed, and won intermediate prizes, but to win the grand prize, they were motivated to pool their insights. The winning algorithm, developed by a composite team, improved the predictive accuracy by 10.09%. A phenomenally cheap piece of R&D for Netflix, a common “Big Data” set as infrastructure, and hacker teams fluidly competing and collaborating. An alliance of the very large and the very small.

More recently Orange, the French telecommunications company, released a data set of mobile phone usage in Ivory Coast, where the company is the sole local carrier. The data recorded the usage patterns of some 50,000 randomly selected individuals over a five-month period, deeply anonymized. It showed how cellphone users moved from place to place, and who (by location) spoke to whom. The idea was to allow researchers simply to see what they could find with such an unusually rich data set. One of the most interesting projects was an analysis by some researchers with IBM⁷ of travel patterns in Abidjan, the largest city. They used the cellphone data to understand where people originated and ended their daily commutes. This enabled them to re-optimize the bus routes in the city, potentially cutting the average commuting time by 10% without adding any buses. Another powerful application could be in public

health, where patterns of physical mobility predict the spread of epidemics, and patterns of communication can be tapped in propaganda campaigns to help combat disease. This promises a revolution in public health.

In all probability, Orange alone could never have identified these questions, still less solved them: they are a phone company. But the value of the data is bigger than the industry in which it originated, and by opening the data to investigation by all-comers, Orange is pioneering a new way of thinking about their business. Perhaps at some point in the future, phone companies will give away telephony and make their profits from the data: it sounds far-fetched, but so did free research services before the advent of Google. Orange are right to experiment: in the world of Big Data, the insights that the data will yield are unlikely to be knowable before the fact, still less will they be most apparent to the institution that happens to put the data together.

Too Small: Building Data Infrastructure

Big Data scales beyond the confines of the traditional business model in the operation of physical facilities; so companies are outsourcing data processing tasks to the providers of “cloud computing.” Cloud providers such as Amazon Web Services enjoy economies of mass relative to their customers. Most departmental servers running one or two applications in the corporate environment achieve only 10-15% utilization, because of the need to provision capacity to accommodate the occasional peaks. Amazon can achieve higher utilization by exploiting the Law of Large Numbers: as long as demand fluctuations are somewhat independent, their sum is proportionately less volatile. Thus Netflix can efficiently serve its movies from Amazon facilities because its peak times—evenings—are out-of-sync with the peak times for many of Amazon’s other corporate customers: work hours. Equally important, *managing* such facilities is a specialized skill: an increasingly sophisticated “core competence” that typical corporations may lack. Specialists can manage uptime, back-up, disaster recovery, upgrades, and patches with greater sophistication than can most end-users. They can respond faster to security threats. The cloud provider thus focuses on the classic virtues of general-purpose infrastructure: reliability, ubiquity and efficiency. Customers save money, but more important, gain flexibility. They can mobilize resources, scale up processes, even deploy entire new businesses, in a matter of hours instead of weeks. Flexibility and cheap adaptation are enabled by breaking a traditional value chain into its components and managing the scale-sensitive pieces in a separate organization.

But this story is not confined to facilities: the same logic applies to the data itself. Since Big Data opens the possibility of much larger data sets and far more sophisticated analytics, this can open new opportunities for competitive advantage.

In 1994 Tesco, the UK grocery retailer, piloted a new loyalty card called Clubcard. They hired a husband-and-wife team, Clive Humby and Edwina Dunn, both mathematicians, to do something revolutionary: understand customer behavior using what we would now call “Big Data.” Clubcard gave Tesco granular transaction data, by SKU, checkout location, customer, and shopping trip. Dunn and Humby mapped the Tesco product range across about fifty abstract dimensions: size, price-point, color, sweet-salty, and so forth. They then looked at the baskets of goods that families purchased to establish correlations among these dimensions. Purchase of “marker products” revealed households’ previously invisible segmentation variables such as budget consciousness, status anxiety, and vegetarianism. Plus segmentation variables that nobody could explain, and nobody *needed* to explain: in the world of Big Data correlation suffices. Tesco then used these correlations to identify non-obvious customer predilections, to identify product pairs that are variously substitutes or complements, and to promote across categories.

The results were spectacular. Redemption rates on promotional coupons reached 20% (compared with 1% for the industry at large).⁸ Tesco saved about \$350 million by targeting promotions more efficiently. And, propelled largely by Clubcard, Tesco overtook Sainsbury’s to become the leading grocer in the United Kingdom.

Big Data scales beyond the confines of the traditional business model in the operation of physical facilities; so companies are outsourcing data processing tasks to the providers of “cloud computing”

For some years, rival Sainsbury’s struggled to find a response. Tesco’s lead in scale and cumulative experience appeared insurmountable. Sainsbury’s eventual and bold move was to *outflank* Tesco by opening Nectar, their new loyalty card, to other retailers. Nectar was launched in conjunction with department store Debenhams, oil giant BP, and credit card company Barclaycard, and managed by a neutral party, a company called Loyalty Management Group. Others have joined since. Nectar users get points to spend at more retail outlets, and Nectar gains both scale and scope in its user data. Sufficient scope might compensate for the initial disadvantage in scale and experience. But note the critical principle: in the era of Big Data economies of mass can extend beyond the boundaries of the traditional business definition; and so value and advantage can be created in new institutions that pool the data.

The same logic is likely to play out on a much larger scale in genomic medicine. Big Data techniques will be used to see fine-grained patterns among individuals’ genomic data, medical history, symptoms, protocols, outcomes, real-time data from bodily sensors, and ambient data from the environment.

Medicine will advance by decoding *immense, linked, cheap, noisy* data sets, instead of the *small, siloed, expensive, clean, and proprietary* data sets generated by hospital records, clinical trials, and laboratory experiments. By accessing such databases,

In the era of Big Data economies of mass can extend beyond the boundaries of the traditional business definition; value and advantage can be created in new institutions that pool the data

practitioners and even patients can become researchers, and evidence-based best-practice can be faster diffused across medical communities.

But an awkward question arises: how can such data be melded when providers, insurers, device companies, pharma companies, Google, patients, and governments not only possess different pieces of the data elephant but guard them jealously and compete on

their information advantage? Where pooled data makes sense, how are privacy and patient rights going to be protected? Technology alone cannot solve these problems. The answer—the only possible answer—is architecture. We will need an infrastructure of trusted, neutral data repositories.

These shifts are already happening. Nonprofit organizations are positioning themselves as platforms for the anonymization, curation, and protection of genomic databases. The Million Person Genome Project is up and running, in Beijing. Registries, run by universities and medical associations, are emerging as living repositories for sharing data on evidence-based medicine. New anonymization and encryption technologies reconcile the scientific imperative to share with the personal right to privacy. Building a shared data infrastructure will be one of the signal strategic challenges of the next decade for the healthcare industry and for policymakers.

The Manager's Agenda

It goes without saying that the most immediate agenda with respect to Big Data is operational. People responsible for market research, process engineering, pricing, risk, logistics, and other complex functions need to master an entirely new set of statistical techniques. Highly numerate analysts trained as recently as ten years ago are waking to the discovery that their skills are obsolete. IT departments need to master data processing on an entirely different scale, and frequently in real time rather than offline batch processing. Non-specialist managers need to understand enough about the possibilities and pitfalls of Big Data to translate its output into practical business benefits. Data visualization is emerging as a critical interface between the specialist and the non-specialist. But every company, eventually, will get there: like the

transition from paper spreadsheets to Excel, the new capabilities will simply be “table stakes,” not a source of sustainable competitive advantage.

The bigger issue is the potential for Big Data to “disrupt,” both as a threat and an opportunity. Deconstruction and polarization of economies of mass are the two key vectors of attack. Deconstruction allows an insurgent to pick off a vulnerable sliver of another company’s value chain, even in apparently unrelated businesses. “Negative” polarization of economies of mass allows small companies, maybe even communities of unpaid individuals, to swarm over a task in ways that corporations cannot easily replicate. “Positive” polarization of economies of mass allows corporations with really large data sets to force their way into new businesses, often giving away the product or service just to access even more data. In an alliance of the big with the small, these corporations often expose some of their data to communities, thus attacking the traditional business model from both sides.

In response, the incumbent corporation has to do precisely these things to itself. It needs to deconstruct its own value chains, open some of its own resources to the energies of communities, and, by one means or another, push some of its resources over a much higher threshold of critical mass. This is true whether the purpose is attack or defense. It may require redrawing business boundaries and redefining relations with customers and suppliers. It may require outsourcing functions previously regarded as “core.” In some functions it will require a radical decentralization or devolution of authority, perhaps beyond the corporate boundary. In others it will require a radical centralization of resources. The key point—indeed the key corollary of deconstruction and polarization—is that these apparently contradictory strategies are mutually complementary.

As Big Data reshapes business, it will transform two fundamental aspects: internal organization, and industry architecture.

Organizationally, Big Data impels corporations to consolidate databases in order to achieve internal economies of mass. They need to establish a “single point of truth” in real time. This can be an immense challenge, because information on the same customer can be locked in different product lines and different channels. Most corporations cannot connect their online and offline data seamlessly. Rebuilding legacy databases from scratch is infeasible, so managers need to craft a migration path by which investments in a new, more functional architecture can pay for themselves as they are implemented. The legacy data warehouse needs to be shut down, but in stages. The financial case for doing this can appear unimpressive, but it must be evaluated *strategically*. Otherwise, a new entrant, with no legacy, will enjoy an immense advantage.

Big Data impels corporations to consolidate databases in order to achieve internal economies of mass

Conversely the analytical skills to query that integrated database, to find those “big insights,” need ultimately to be decentralized into the business units. That will take time, since today those skills are in very short supply, and must be rationed. Corporations need to develop explicit plans to manage this evolution.

The implications of Big Data for industry architecture are all about tapping the superior capabilities of other players. It may require outsourcing innovation to small contributors, especially customers, by exposing APIs and proprietary databases. It may require outsourcing processing and facilities management to a cloud provider that enjoys superior economies of scale and experience. It might involve investing in data partnerships to achieve critical mass collectively that would be infeasible severally. In every case the definition of the business is being changed to accommodate the evolution of competitive advantage beyond the bounds of the traditional business model.

There is one final issue that is really beyond the scope of this chapter, but whose importance cannot be over-emphasized: data rights. It is profoundly ambiguous in most business contexts who “owns” personal data and what rights they have to use it. In principle there is a contract between the data subject and the data user that governs this question. But in practice it is pretty meaningless: data subjects do not read the contracts, have little choice but to sign, and do not know how their data is actually being used. If the terms of data exchange were tightened, as some policymakers have proposed, then the properly open-ended nature of Big Data exploration would be stymied. It is unlikely that these legal and perceptual ambiguities will be cleanly resolved in the next few years. In the interim, corporate (and governmental) use of personal data will depend critically on the *context* in which the data is gathered and used, and on the degree of *trust* enjoyed by the data-using organization. Establishing that context, and building that trust, will be fundamental challenges. Ultimately the legitimacy with which corporations use their data, in the eyes of their customers and the eyes of society, will constrain the rate at which the Big Data revolution transforms our world.

Notes

1

<<http://www.businessinsider.com/the-future-of-mobile-slide-deck-2013-3?op=1>>

2

<<http://newsfeed.time.com/2013/03/25/more-people-have-cell-phones-than-toilets-u-n-study-shows>>

3

<<http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook>>

4

<<http://research.gigaom.com/report/a-near-term-outlook-for-big-data>>

5

<<http://www.sciencedaily.com/releases/2013/05/130522085217.htm>>

6

<http://www.macobserver.com/tmo/death_knell>

7

M. Berlingerio, F. Calabrese, G. Di Lorenzo, R. Nair, F. Pinelli and M.L. Sbdio, "AllAboard: A System for Exploring Urban Mobility and Optimizing Public Transport Using Cellphone Data," *Machine Learning and Knowledge in Databases*, LNCS 8190 (2013), 663-66.

8

<<http://www.information-age.com/channels/information-management/it-case-studies/277256/getting-relevant.thtml>>



OpenMind Channel



Article

- From Deconstruction to Big Data:
How Technology is Reshaping the Corporation

About the Author

Philip Evans

Related Articles

- Big Data and the Future of Business
- Who Owns Big Data?
- The Internet and Business

Read the Full Book

- Reinventing the Company in the Digital Age

Other Books



Share

