Ubiquity Symposium

Big Data

Technology and Business Challenges of Big Data in the Digital Economy *by Dave Penkler*

Editor's Introduction

The early digital economy during the dot-com days of internet commerce successfully faced its first big data challenges of click-stream analysis with map-reduce technology. Since then the digital economy has been becoming much more pervasive. As the digital economy evolves, looking to benefit from its burgeoning big data assets, an important technical-business challenge is emerging: How to acquire, store, access, and exploit the data at a cost that is lower than the incremental revenue or GDP that its exploitation generates. Especially now that efficiency increases, which lasted for 50 years thanks to improvements in semiconductor manufacturing, is slowing and coming to an end.

Ubiquity Symposium

Big Data

Technology and Business Challenges of Big Data in the Digital Economy

by Dave Penkler

The almost exponential growth in digital data can be attributed to technology and business factors. On the technology side, there have been three main contributors: devices, networks, and storage.

The evolution of sensor technology and embedded computing has led to a dramatic increase in the number of devices and the volume of data they generate. Such devices include PCs, tablets, smartphones, wearables, environmental sensors, and surveillance cameras.

Advanced coding and modulation techniques implemented in modern digital signal processors has enabled spectrum efficient radio network access in and around the planet. Satellite networks, the evolving cellular networks, low-power wide area networks such as LoRa or narrowband IOT, Wi-Fi and bluetooth low energy (BLE) networks are some examples. Improvements in optical transmission and modulation technologies have provided the bandwidth to move data quickly and efficiently around the globe.

Finally the impressive progress the storage industry has shown in increasing the density of stored data and reducing the cost in terms of Euros/gigabyte over the years has fueled the growth of stored data.

On the business side the drivers to collect data have been:

- Monetization of digital assets
- Optimization of operations
- Improvement of services and customer experience
- Exploitation of business intelligence

Companies are retaining logs from IT systems, applications, and networks over long periods of time in order to exploit the residual value they may have. User, partner, and employee interactions with devices and applications are being held in data stores to improve and optimize business outcomes. Network operators in many countries must keep call records by law, but these are also used for business intelligence to predict user churn for example. Data from monitoring user activity, movement, and the sites they access provides additional insight into customer behavior and in certain cases this data can also be monetized. The tendency of companies to hold on to as much data for as long as possible, in the hope it may be useful, could be called "the squirrel effect," which is substantially contributing to the growth of corporate big data.

So what is big data then? The term "big data" means different things to different people. For some it is synonymous with business intelligence or analytics, others look at it through a technology lens seeing it as data handled by certain tools. A traditional view is data characterized by the three Vs: large **volume**, it accumulates with high **velocity** and is of great **variety**. The definition of big data as large data sets that cannot be efficiently handled with traditional IT tools seems to be the most appropriate for an IT practitioner. In other words, it is large data that is difficult to move, store, analyze, and exploit with the current networks, file systems, databases, and tools.

The early digital economy during the dot-com days of internet commerce successfully faced its first big data challenges of click-stream analysis with map-reduce technology. Since then, the digital economy has been becoming much more pervasive. The dematerialization of money and content is well underway with bitcoins, digital books, photos, audio, movies, art and services. Expanding the scope one can also consider robots as part of the digital economy. For example, there are software robots running on personal devices or in the cloud that one can talk to do simple things like provide information or open the blinds. The software robots that conduct trading at lightning speed, and other software agents responding to events and conditions on behalf of businesses and governments, are all actors in this digital network mediated economy. Expanding the scope further, it could include the mechanical robots used for manufacturing in fully digitally controlled factories where no human intervention is required for routine operation. Into this category one could include also the drones used for space, undersea, or land exploration or even drones used in economically motivated remote military operations. In the long term, finding an equitable and sustainable economic balance between consumers and producers in the digital economy will probably be its biggest challenge.

As the digital economy evolves, an important technical-business challenge is emerging: How to acquire, store, access and exploit the data at a cost that is lower than the incremental revenue or GDP that its exploitation generates. Especially now that efficiency increases, which lasted for 50 years thanks to improvements in semiconductor manufacturing, is slowing and coming to an end.

Typical large data analysis for supply chain and logistic operations requires the construction of huge graphs that can be analyzed very rapidly in order to respond to or anticipate ever changing conditions. The nodes of the graph represent the attributes of the entities involved. Its arcs are the spatial, temporal, and other relationships of interest that exist between the nodes. For an airline logistics operation the nodes would include passengers, aircraft, airports, crews, and so forth; the arcs would be routes, schedules, resource dependencies, etc. Graph processing algorithms access memory in small non-contiguous chunks spending a few cycles on each node before following one of the relationship arcs to the next node. The achievable processing parallelism varies dynamically due to the different degrees of fan-out and fan-in of the relationships being explored. Typical massively horizontally scalable analytics platforms, such as Hadoop, are not up to the task since the relationship arcs cross machine boundaries causing a lot of horizontal interprocess messaging overhead. What is needed to avoid this is a large shared memory machine that allows any one of its many thousands of processing elements to access any part of the graph directly. Further efficiencies in performance and energy can be gained by reducing data movement between the different levels of the traditional storage hierarchy. This can be achieved by reducing the memory hierarchy to a single persistence layer of byte-addressable non-volatile memory (see Figure 1).

Figure 1. Collapsing the memory hierarchy

In other words the DRAM, PCIe attached NVRAM, and SATA or network attached storage layers need to be collapsed into a single level that can hold the whole graph in a single address space, without incurring a substantial increase in cost [1]. To meet the required performance the maximum two-way memory access latency must be kept within half a microsecond (500 nanoseconds). This poses a number of challenges: Firstly the cost of memory must drop substantially below that of current DRAM prices. Secondly, in order to meet the performance and energy requirements an optical communications fabric will be required connecting chips, boards, chassis, and racks. However, the overall size of the machine will be constrained by the speed of light in an optical fibre. Fabric switches are used to reduce the number of point-to-point links. The propagation speed of light in an optical fibre link is approximately five nanoseconds per metre. So, ignoring the latency incurred by the memory modules and fabric switches, for a 250 nanosecond one-way processor to memory latency the physical machine would have to fit into a circle or sphere of 50 metres in diameter or less depending on other sources of latency. In order to effectively cool such a dense machine highly efficient non-volatile memory will be essential. The jury is still out on which non-volatile byte-addressable memory technology will be able to replace the mechanical and solid state block addressable storage technologies. Among the candidates are the different resistive random access memories (ReRAMs) based on phase change materials, conductive metal oxides, or conductive bridging metals. Spin Transfer Torque RAMs, while very fast and reliable, will likely not achieve

the required density in time. Increasing the size of the machine will increase the worst case memory access latency. To mitigate this effect will require increasing the capacity for tracking outstanding memory transactions in the memory controllers and memory fabric switches. Until the non-volatile memories (NVMs) reach the required performance, the computational working sets will be held close the processing elements using high bandwidth DRAM modules using the fabric attached NVMs as a buffer and backing store. Determining where along this trade-off line the point of diminishing returns might be, has been a long standing research and development question, especially in the high-performance computing field.

While not directly in the mainstream of the digital economy, the analytics of big data in science and research is an important tributary. In high performance computing (HPC) there is big data for HPC where for example the extremely voluminous execution traces of high performance computing applications need to be analysed in order to optimize data placement in memory to maximize performance. Then there is HPC for big data as in fine-grained weather simulations [2]. A machine such as that mentioned earlier could be used for molecular dynamics simulations, which enable the design of quieter, more efficient and environment friendly turbo-fan engines for passenger aircraft. This is achieved by the simulating airflow and combustion reactions at the molecular level. It can be assumed the technologies being developed for big data in science and research will be adapted and adopted downstream in the digital economy infrastructure.

Looking forward one might ask whether big data in the digital economy will improve or deteriorate the state of the world. The answer to this question obviously depends on the perspective with which the world is viewed. Certainly from the economic point of view of the large corporations things will improve in the short term. Big data analytics brings greater efficiencies while broadening and deepening commercial reach. For the semi-skilled workers, both in developed and less-developed countries, the outlook is somewhat grimmer. Indiscriminate monitoring and software control of almost all aspects of their work and living environment including manipulation of their behavior using personality profiles from social media big data will lead to a decrease of their relevance in the global economy as well as disenfranchisement in society. Corporations will deploy their new fully automated factories in countries with developed digital infrastructures and favorable fiscal legislation rather than in countries with large low cost labor forces.

From the perspective of security and privacy, big data will worsen the situation. To address the security issue one needs to establish trust regarding the identity of the objects, consumers, and producers in the digital economy. The concomitant loss of privacy would exacerbate the disenfranchisement of lower income workers since the precision with which the analytics can

pinpoint individuals and the influence their behavior will be increased. On the other hand the effective anonymization of data with respect to individuals and infrastructure is complex and not well understood [3]. Furthermore, dealing with cybersecurity challenges is already becoming a major cost factor for IT service providers. It is making software and software development more complex and as a result also more error prone, especially in the so called “internet of things.” As the scope of the digital economy expands the risks and costs will increase. Time and money spent on security does not create value for citizens and consumers.

Environmental considerations add another perspective to this question. If, in war and business, knowledge is power and having pertinent knowledge ahead of the enemy or competition can mean having a decisive advantage, then it can be expected that big data and its analysis will consume ever more energy. The keen eye of sophisticated analytics applied to seismic and remote sensing data in fossil fuel exploration is not likely to improve the environment for many of the living organisms on this planet. Unsupervised learning that is widely used in machine learning is low in human capital but very high in computing, network, storage, and cooling energy costs. Its brute force approach requires crossing huge amounts of data with itself in many different combinations. In order to sustain the growth of the economy, the choice of digital media presented to consumers is shaped by the profiling of their behaviors and aspirations and used to create demand in already saturated markets. The feigning of the scarcity of products due to the implied avidity of other online consumers is a common technique used to inveigle the consumer to spend their money on non-essential goods sooner. The overproduction afforded by automation associated with the highly sophisticated demand generation from advertising powered by big data and its analysis is having a deleterious impact on the environment.

A rough estimate of the energy consumed per social media interaction or web search request is on the order of six kilo joules, which is the equivalent of the energy needed to boil a cup of water. One may ask what proportion of this energy is expended on the computation behind the placement of advertisements and the ranking of results or the selection of the feeds. A recent realization among some sustainability conscious consumers might serve as an example of a way out of this somewhat dystopian outlook. These modern denizens of the digital economy, who unplug their smartphone chargers when not in use, are realizing their inadvertent use of cloud services is contributing to climate change. They are taking the initiative and determining, albeit in a small way, the broader outcomes. By exercising their choice, based on an informed conscious compromise between their personal needs and their concern for the environment

they are freeing themselves from the bondage of the potential consequences of big data analytics discussed above.

The question of how analytics of big data in the pharmaceutical and agricultural industries is affecting people's health should also be examined more carefully. Software tools for molecular modelling and genetic analysis used to design pharmaceutical and genetically modified seed products are extremely complex and require enormous computational resources. The current limitations of the tools together with the dearth of computational resources leave room for doubt about how well the direct and indirect side effects that these products could have on human organisms are understood.

How can the majority of the participants in the digital economy be re-enfranchised? One way is to promote the proactive symbiosis between humans and their digitally mediated environment. An environment where each human participant can in fact "program" behavior of the mediation as opposed to simply being subjected to the rules that have been determined by machine learning and the highly skilled programmers in the employ of the potentates. Another is the targeted collection and opening up of big data by governments and institutions for the use of individuals and small businesses that would otherwise not have the means. Open data repositories exist today but they are small, patchy, and not well curated. Researchers have expressed concern that they cannot verify the results of big data analyses published by colleagues who are affiliates of large private companies.

Opening up big data is problematic but not impossible. One major issue is of course security and privacy, since the open data could also be used by many for nefarious purposes. But lessons can be learnt from the non-digital world—the public transport environment for example. There are roads that determine the paths cars and trucks can take. Vehicles must have roadworthy certificates that need to be renewed periodically, and drivers must have licenses that can be revoked in case the rules of the road are not respected. Roadworthiness and respect of the rules of the road are enforced by legislation, police, and surveillance equipment. In the same way open, big data analytics environments operated by non-commercial national or supranational institutions can restrict and control their use with modern cloud computing techniques. By the very nature of big data it is not amenable to be downloaded for local processing. For efficiency and security reasons it needs to be processed in situ. The cost of the public infrastructure to support a big data analytics platform for environmental monitoring and traffic data in a big city, for example, would not exceed that of 100 kilometres of highway. As the scope of the digital economy expands it is to be hoped that those whose remit it is to look after the interest of the citizens will take the initiative to ensure their right to self-determination, including of course the citizens themselves.

Disclaimer: The opinions expressed herein are the personal opinions of the author and do not necessarily represent those of the employer.

References

- [1] Bresniker, K. et al. Adapting to Thrive in a New Economy if Memory Abundance. *IEEE Computer* 84, 12 (Dec. 2015), 44-53.
- [2] Miyoshi, T. et al. "Big Data Assimilation" Toward Post-Petascale Severe Weather Prediction: An overview and progress. *Proceedings of the IEEE* 104, 1 (Nov. 2016), 2155-2179.
- [3] Lepri, B. et al. [The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-Making for Social Good](#). [preprint version]. To appear in "Transparent Data Mining for Big and Small Data," Studies in Big Data Series, Springer.

Biography

Dave Penkler is a technologist in Hewlett-Packard Enterprise's Communications and Media Solutions business, where he is responsible for forward-looking technology in the Internet of Things for service providers. Specific areas include low power wide area networking and edge computing. His other interests include the application of silicon photonics in large-scale datacenter networks, opensource, LISP and APL. Dave has more than 35 years experience in designing and programming operating systems, networking and telecommunications systems. He is an HPE Fellow and holds a B.Sc. in mathematics and computer science from the University of the Witwatersrand Johannesburg.

DOI: 10.1145/3158341