Industrial Management & Data Systems

Big Data promises value: is hardware technology taken onboard?

Wasim Ahmad Bhat S.M.K. Quadri

Big Data promises value: is hardware technology taken onboard?

Wasim Ahmad Bhat and S.M.K. Quadri
Department of Computer Sciences, University of Kashmir, Srinagar, India

Received 29 April 2015
Revised 8 September 2015
13 September 2015
Accepted 14 September 2015

Abstract

Purpose – The purpose of this paper is to explore the challenges posed by Big Data to current trends in computation, networking and storage technology at various stages of Big Data analysis. The work aims to bridge the gap between theory and practice, and highlight the areas of potential research.
Design/methodology/approach – The study employs a systematic and critical review of the relevant literature to explore the challenges posed by Big Data to hardware technology, and assess the worthiness of hardware technology at various stages of Big Data analysis. Online computer-databases were searched to identify the literature relevant to: Big Data requirements and challenges; and evolution and current trends of hardware technology.
Findings – The findings reveal that even though current hardware technology has not evolved with the motivation to support Big Data analysis, it significantly supports Big Data analysis at all stages. However, they also point toward some important shortcomings and challenges of current technology trends. These include: lack of intelligent Big Data sources; need for scalable real-time analysis capability; lack of support (in networks) for latency-bound applications; need for necessary augmentation (in network support) for peer-to-peer networks; and rethinking on cost-effective high-performance storage subsystem.
Research limitations/implications – The study suggests that a lot of research is yet to be done in hardware technology, if full potential of Big Data is to be unlocked.
Practical implications – The study suggests that practitioners need to meticulously choose the hardware infrastructure for Big Data considering the limitations of technology.
Originality/value – This research arms industry, enterprises and organizations with the concise and comprehensive technical-knowledge about the capability of current hardware technology trends in solving Big Data problems. It also highlights the areas of potential research and immediate attention which researchers can exploit to explore new ideas and existing practices.

Keywords Big Data, Networking, Storage, Knowledge economy, Microprocessor, Technology trends
Paper type Research paper

1. Introduction

Earlier technology only provided a platform for processing data to yield information in order to improve the performance of existing processes, businesses and activities. With the advancement and proliferation of technology, the digital data nowadays comes from variety of sources (countless sensors, innumerous web applications, growing handheld devices and so on), in various forms (text, images and videos), in huge volumes and with high velocity. Big Data is the term used to describe this voluminous, heterogeneous and frequent generation of semantically unstructured data (Laney, 2001). Big Data is large in volume, complex in structure and aggressive in its production. Information and communication technology has always had a positive

The authors thank their colleagues from the University of Kashmir, especially Dr Ameen Parray, who offered his invaluable expertise. The authors would also like to thank the anonymous reviewers for their critical assessment of the manuscript, detailed insights, and valuable and encouraging comments.

Industrial Management & Data Systems
Vol. 115 No. 9, 2015
pp. 1577-1595

0263-5577
DOI 10.1108/IMDS-04-2015-0160

impact on innovation activity (Kmieciak *et al.*, 2012), performance (Bayo-Moriones *et al.*, 2013) and value (Ong and Chen, 2014) of an enterprise; Big Data being no exception. Big Data provides an opportunity to academics, industry and organizations to better understand a process or a phenomenon. If proper analysis of Big Data is performed and correct correlations are drawn, Big Data promises generation of new ideas, innovations, new products, higher productivity and profitability. Research has shown that knowledge creation has the potential to act as a catalyst for innovation (Begona Lloria and Peris-Ortiz, 2014) and Big Data not only promises creation of new knowledge but also new kinds of knowledge, on which an entirely new economy can be founded (Haynes and NGuyen, 2014). Big Data also promises entirely new classes of economic activities built on insights, and the value derived from it.

Immense technology advancement and its intrusion into every aspect of our life are the basic reasons for creation and growth of Big Data. If this proliferation and penetration of technology continues, richer and heavier data sets would be created. On the other hand, if advancements in technology support proper, diverse and timely analysis of such data sets, stronger and stable knowledge economy would be created. This implies that for a sustainable knowledge economy to be an outcome of Big Data, technology must play its part effectively at all stages of Big Data. Although the benefits of Big Data are being unanimously envisioned across the globe, the challenges are still being discussed and their solutions are yet to be finalized. One such challenge is faced by hardware technology to support Big Data creation, growth, communication and analysis for knowledge creation. National and international projects such as the Large Hadron Collider at CERN, and such many other, are frequently cited for the way they will challenge the state of the art in three main aspects of hardware technology, i.e., computation, networking and storage (Lynch, 2008). The problem is exaggerated by the fact that current technology trends in computation, networking and storage have not evolved keeping in view Big Data, though same is being employed in Big Data from its generation to analysis. Therefore, allowing technological gaps to creep in at all stages of Big Data; rendering Big Data incomplete and its analysis inappropriate.

Academics and practitioners alike, therefore, face several questions: what are the challenges posed by Big Data to current trends in computation, networking and storage technology at various stages of Big Data analysis? How effectively does current technology trends in computation, networking and storage support Big Data at various stages? Are there any technological gaps between Big Data requirements and current trends in computation, networking and storage technology? What are the areas of potential research in computation, networking and storage technology with respect to Big Data? Many studies in past have analyzed the general challenges posed by Big Data to technology. Also, many focussed-studies have analyzed the challenges posed by Big Data to specific areas of technology like security (Tankard, 2012), usability (Jianzhong and Xianmin, 2013), privacy (Kshetri, 2014), data management (Little, 2012), cloud (Agrawal *et al.*, 2011) and so on. Unfortunately, studies performing assessment of effectiveness of technology employed, exploration of existing technological gaps and identification of areas of potential research with respect to computation, networking and storage technology (collectively) is yet to be done. Therefore, leaving aforementioned questions yet to be answered. Answering such questions would bridge the gap between theory and practice, and would point toward areas of potential research, besides significantly contributing to knowledge base and pinpointing stage-specific effective-technology. Bearing this research in mind, the aim of this paper is to analyze the challenges posed by Big Data to current technology trends in computation, networking and storage. It addresses following research questions:

RQ1. What are the challenges posed by Big Data to current trends in computation, networking and storage technology at various stages of Big Data analysis?

From this, two main sub-questions emerge:

RQ2. What stages of Big Data analysis are effectively supported by current trends in computation, networking and storage technology?

RQ3. What are the technological gaps, and areas of potential research, in computation, networking and storage technology with respect to Big Data analysis?

The reminder of the paper is organized as follows: Section 2 introduces the research methodology used for the study which employs literature review approach. Section 3 describes trends in computation technology and challenges posed by Big Data at data generation and analysis. Section 4 discusses trends in networking technology and challenges posed by Big Data at data transportation and analysis. Section 5 describes trends in storage technology and challenges posed by Big Data. The final section discusses and summarizes contributions to research and practice.

2. Research methodology

Relevant research concerning Big Data challenges and current trends in hardware technology was identified by searching the online computer-databases for primary research material. These online databases included ACM Digital Library, IEEEXplore, Web of Science, Sciencedirect, SpringerLink, Arxiv.org, Inspec, Scirus, Scopus, CiteSeerx and such many other. They were searched for publication from 2000 through present. In order to ensure that relevant studies are not missed, the search terms remained broad. The search terms were categorized into two: first, terms which were meant to identify studies investigating Big Data challenges and requirements; and second, terms which were meant to identify studies investigating evolution and current trends in hardware technology. In former case, the search terms included “Big Data” AND “challenges OR requirements OR hardware challenges OR hardware requirements OR technology challenges OR technology requirements.” In latter case, the search terms included “hardware OR technology” AND “evolution OR trends OR future.” Many alternative keywords with similar meanings were created and were employed to identify the relevant studies. Furthermore, a comprehensive search using same keywords was made for internet resources which yielded white-papers, technical-reports and so on. Due to usage of broad and alternative keywords, the process yielded 227 related studies.

The next step was a detailed examination of papers, and at this point such studies were excluded in which: first, Big Data challenges described were either insufficient or not applicable; or second, evolution of hardware technology was incomplete; or third, current and prospective trends in technology were either obsolete or not applicable; or fourth, accuracy and credibility of the content was questionable. The bibliography of examined papers was also checked for further identification of relevant studies (which might have been missed earlier). Based on the filtering criteria, following number of studies were excluded for being either not applicable or incomplete: 36 studies related to Big Data challenges, 24 studies related to microprocessor technology, 32 studies related to networking technology and 18 studies related to storage technology. Furthermore, 18 internet resources and 12 scholarly articles were discarded based on the content accuracy and credibility criteria. Also, 13 studies were excluded for being out-of-date. Finally, a total of 113 studies were included for the study. Figure 1 shows the outcome of filtering criteria employed to identify the most relevant literature for the study.

For a systematic and critical review of literature, articles were first previewed for thematic organization of relevant literature. After that, each article was read, questioned and summarized. This technique is effective, if a large number of publications are being reviewed. It facilitates easy identification and retrieval of material, besides keeping the reviewer focussed and consistent. For each article, the issues highlighted and the solutions proposed were recorded. With respect to literature related to Big Data challenges, it was observed that "opportunities" and "challenges" are being pronounced together in same breath. This indicates that even though the wide range of applications of Big Data are promising, the challenges posed by it exist, and, therefore, need attention. Furthermore, the studies revealed that the challenges posed by Big Data to technology exist at all the three main stages of Big Data analysis: source; transportation; and sink. On the contrary, in case hardware technology trends, it was observed that the evolution and current trends in three main aspects of hardware technology (computation, networking and storage) had paid little attention to Big Data. The studies revealed that technology had evolved with various parameters into consideration except Big Data. However, it was also observed that with the growing popularity of Big Data, the technology trend is shifting to facilitate Big Data at all stages of analysis. Perspective repetition, and repetition of research approach was also observed in many studies belonging to both categories of the included literature. Consequently, many hardware trends, and challenges posed by Big Data emerged which were categorized into three main categories: computation; networking; and storage. In each category, some key themes were identified which include: trends at source/transportation; challenges at source/transportation; trends at sink; and challenges at sink. Under "trend" category, the evolution of technology to its current trends for each category was identified.

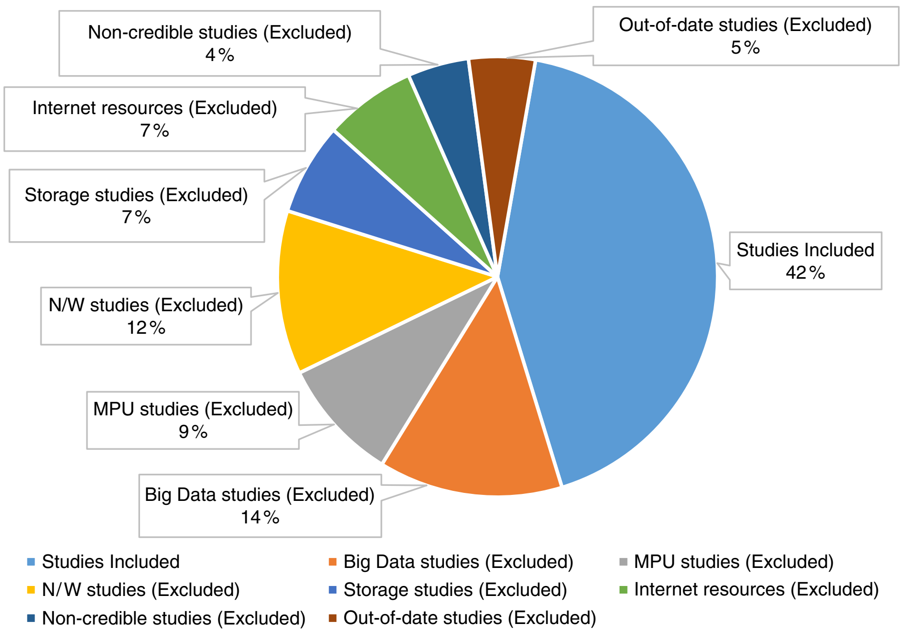

Figure 1.
Identification of most relevant literature included in the study after filtering

It was mapped with Big Data challenges, in an attempt to understand how effectively the trends in the technology facilitate Big Data. Under "challenges" category, Big Data challenges which were not satisfied by current trends in the technology were identified.

3. Computation technology

In general at Big Data source, a digital processing unit is employed to digitize the raw data gathered. As an example, a sensor that monitors the heart rate of a patient in hospital uses a digital processor to convert heart beats into appropriate digital data. On the other side of the spectrum, Big Data promises value, if and only, correct and timely analysis is performed which also requires services of a digital processor. However, the computation requirements at Big Data source and sink are totally different. This section discusses the current trends in computation technology applicable to Big Data generation and analysis, and highlights its various shortcomings and challenges.

3.1 Trend in data generation

Moore's (2006) law has governed microprocessor technology advancements for past several decades by predicting that the number of transistors on a chip doubles after every two years. As a consequence, roughly after every 18 months, the transistor density has doubled, fabrication cost has reduced, performance has increased by 40 percent and chip size has reduced by 30 percent (Pollack, 1999). Availability of sufficient computing power, small size, low power consumption and low cost is primarily responsible for lengthwise and breadthwise penetration of digital technologies into every aspect of our life. The wide domain applications of such digital technologies has been embraced by all organizations and people from different walks of life, leading to generation of voluminous amounts of digital data. Of 7 billion things connected to the internet in 2010, 5 billion were not computers. These other devices range from internet enabled cell phones, networked entertainment and gaming devices to automobiles, building automation systems, smart meters and thermostats, medical electronics, and industrial controllers (Vesset *et al.*, 2012). For the same reasons mentioned above, this digital data were heterogeneous in nature and comprises of textual data, images and videos. This semantically unstructured data is growing in size at 62 percent annually in contrast to semantically structured data growing at 22 percent (Mukherjee and Krishnamurthy, 2012). Furthermore, due to growth and proliferation of digital technologies, digital data has been exploding since year 2000 when it accounted for roughly 25 percent of all the information stored, and by 2007, its share increased to 94 percent (Mukherjee and Krishnamurthy, 2012). The heterogeneous, voluminous and rapid generation of data at source is expected to continue. The reason behind this is that it is inevitable to anticipate more and more intrusion of digital technologies, both in depth and breadth, due to availability of low cost, high-performance and power efficient processors with broad range of application areas.

3.2 Challenges in data generation

Big Data sources can also generate erroneous data. As an example, there is no way of knowing whether the reading of a sensor is correct or not, unless an investigation for a faulty hardware is performed. The problem is that it is very difficult to differentiate between a correct reading and an incorrect one. Therefore, to avoid generation of any such data that would lead to unnecessary communication and storage cost, and possibly incorrect analysis, the sources must be augmented with some intelligence to reduce generation of erroneous data while not compromising on the qualitative data.

Furthermore, sources must also be equipped with the capability to automatically add description of data recorded (e.g. what and how the data were recorded?). Adding this metadata gives some level of semantics to the data at source and is useful during analysis at sink. As an example, in scientific experiments, considerable details regarding specific experimental conditions and procedures may be required to be able to interpret the results correctly, and it is important that such metadata be recorded with observational data.

Currently, Big Data sources are not intelligent enough to either avoid generation of erroneous data or add metadata to the data generated. Arming sources with this capability would certainly help in effective and correct analysis.

3.3 Trend in data analysis

For many decades, processing huge and frequently increasing volumes of data has been a challenging issue. The problem was successfully mitigated in past by processors while following Moore's law to provide the necessary computing power to deal with increasing volumes of data. However, due to limitations imposed by Instruction-Level Parallelism and heat dissipation, performance scaling in single-core microprocessor technology is very limited. Pollack's rule states that if we double the logic in a processor core, then it delivers only 40 percent more performance (Borkar, 2007). In contrast, a multi-core architecture has potential to provide near linear performance improvement with complexity and power. As such, two smaller processor cores, instead of a large monolithic processor core, can potentially provide 70-80 percent more performance while consuming lesser power. Consequently, researchers have investigated multi-core architectures to improve the performance of a microprocessor while keeping power consumption under the control. Keeping the number of cores constant, while doubling the transistor count in a generation, has been lately the trend (Durairajan *et al.*, 2013). However, even with multi-core architecture, the heat dissipation and power consumption still exists (Isci *et al.*, 2006). From small platforms like mobile systems to the largest supercomputers, the power budget is limited. This places power and heat management at the top most position among the parameters that dictate the design of a microprocessor (Hofstee, 2005; Isci *et al.*, 2006). Nevertheless, designs using large-scale parallelism and application-customized heterogeneous cores, or a few large cores and a large number of small cores operating at low frequency and low voltage, have been sought out to be as alternatives to achieve performance and energy efficiency (Hofstee, 2005; Isci *et al.*, 2006).

At the same time, Graphics Processing Unit (GPU) technology has evolved to support general purpose usage, even though it was specifically designed and used for graphics and multimedia applications (Stuart and Owens, 2011). GPUs do not employ caches to mask the latency of memory access and reduce out die bandwidth unlike multi-core processors (Seiler *et al.*, 2008). Instead, they use thread level parallelism to mask memory latency by running other threads when some are stalled waiting for memory. GPUs are becoming prospective artifact for general purpose computing by providing enough computational power to support general purpose workloads at a much lower capital equipment cost and at much higher power efficiency. Specifically, they perform better than multi-core processors at the throughput-oriented workloads that are characteristic of scientific computation and large-scale server applications. Many studies have demonstrated techniques to do general purpose computing with GPUs by mapping general purpose applications onto graphics APIs (Owens *et al.*, 2007). Other studies have shown huge gains in both performance and price-performance of GPUs (Stuart and Owens, 2011).

Apart from multi-core processors and GPUs, High-Performance Computing (HPC) clusters (Buyya, 1999) and Field Programmable Gate Arrays (FPGAs) (Brown *et al.*, 2012; Francis *et al.*, 1992) can also be used for Big Data analysis. HPC clusters (commonly known as supercomputers) are custom designed machines with thousands of cores, while as FPGAs are highly specialized hardware units which are custom-built for specific applications. FPGAs are highly optimized for speed and can be orders of magnitude faster than other processors for certain applications when programed properly (Thomas and Moorby, 2002).

Another class of systems for Big Data analysis includes peer-to-peer networks (Milojicic *et al.*, 2002; Steinmetz and Wehrle, 2005) which involves storage and processing of data over millions of commodity machines connected in a network. It is a decentralized and distributed network architecture where the nodes in the network (known as peers) serve as well as consume resources. Apache Hadoop (Shvachko *et al.*, 2010) is an open source framework for storing and processing large data sets using clusters of commodity hardware. Spark is a next generation paradigm for Big Data processing developed by researchers at the University of California at Berkeley (Madden, 2012). It is an alternative to Hadoop which is designed to overcome the disk I/O limitations and improve the performance of earlier systems. Spark developers have also proposed an entire data processing stack called Berkeley Data Analytics Stack (Franklin, 2013).

The current trend is able to tackle the computation requirements of Big Data analysis by providing different classes of systems for different types of application scenarios. It does so by providing custom designed machines containing thousands of processing cores, custom-built application-specific hardware units, network of commodity machines, and power efficient and high-performance GPUs.

3.4 Challenges in data analysis

For Big Data analysis, scaling for performance is an important factor. In Vertical Scaling, the server contains more processors, more memory and faster hardware. Even though management in case of vertical scaling is easy, yet scaling up could be costly. To put simply, a user has to invest more than what is required currently (in anticipation of future workloads) due to limited space and expansion slots. Multi-core processors, HPCs clusters, GPUs and FPGAs are all examples of Vertical Scaling. Unfortunately, HPC clusters are difficult to scale up after a certain extent while as scaling up Multi-core processors, GPUs and FPGAs becomes extremely costly even though they can produce results in real-time.

In Horizontal Scaling, the workload is distributed across many servers which may be even commodity machines each running its own operating system. As a consequence, the system can be scaled to any limit that too in small increments which lessen the financial burden, but, unfortunately, cannot yield results in real-time. Apache Hadoop is the most prominent example of such a platform. Hadoop is designed to scale up to hundreds (and even thousands of nodes) and is also highly fault tolerant, but is usually slow in delivering results.

It can be safely argued that GPUs and FPGA can handle real-time analysis of Big Data as they have high throughput memory and the data I/O operations are extremely fast, but cannot scale out to large volumes of data. In contrast, Hadoop and Spark can scale out to process huge amounts of data, but usually take more time to deliver the results. As a consequence, there is often a trade-off between real-time analysis requirements and scalability of the data being processed. Unfortunately, there is no such system design that is flexible enough to stand up for any computation requirement of Big Data analysis.

4. Networking technology

Networking occupies the central position in Big Data analysis by facilitating both generation and analysis of Big Data. At one end of the spectrum, it allows the collection of data from sources; while as at the other end, it facilitates Big Data analysis in horizontal scaling paradigm. This section discusses the current trends in networking technology available to facilitate Big Data transportation and analysis in peer-to-peer networks while highlighting the technological gaps.

4.1 Trend in data transportation

Nielsen's (1998) law of internet bandwidth states that a high-end user's internet connection-speed grows by 50 percent per year. This is an outcome of continuous drop in the cost of transmitting a bit over an optical network which decreases by half every nine months as stated by Butters' law (Khan and Pi, 2011). Not only is the bandwidth increasing and the cost of transmission decreasing, but networks are also covering more and more geographical area. Figure 2 shows two classes of wireless networks: Wi-Fi and Mobile. In both classes, in addition to transmission speed, the mobility supported by the network has also been enhancing (Rao and Angelov, 2005).

This trend in technology indicates that both wired and wireless networks are becoming faster, affordable and accessible. This allows digital data to be created and shared across the globe in a faster and affordable way. One can understand why google receives over 2,000,000 search queries every minute, 72 hours of video are added to Youtube every minute, there are 217 new mobile internet users every minute, twitter users send over 100,000 tweets every minute (that is over 140 million per day), companies, brands and organizations receive 34,000 "likes" on social networks every minute, and so on (Juniper Networks, 2012)? This implies that people are more connected to, and dependant on, the widespread and affordable internet.

From 2010 to 2011, the global internet population grew by 6.5 percent (over two billion people) and the vast majority of the world's population has yet to connect (Santosh and Bhandare, 2014). Currently, only one-third of world's population is part of the internet population, but progressively the internet population is increasing (Internetworldstats, 2015). Digital devices like smartphones, tablets, etc., are becoming more and more capable and affordable, and web technologies like social networking, online shopping, etc., are flourishing. And, figures reveal that the transmission cost per bit is dropping, annual internet bandwidth is increasing by 50 percent and wireless

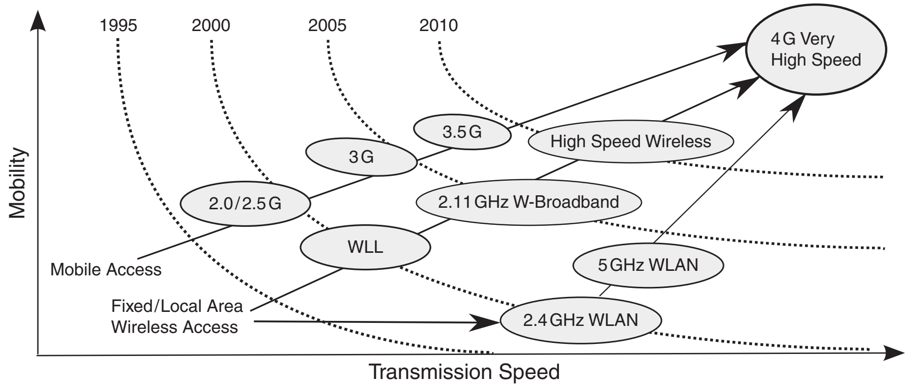

Figure 2.
Advancement in Wi-Fi and mobile networks: mobility and speed

networks are covering more and more geographical area. Therefore, it facilitates the growth of internet population and leads to affordable, speedy and widespread acquisition of Big Data.

4.2 Challenges in facilitating data transportation

The performance of a network is governed by two factors – its bandwidth and latency. Over past 30-40 years, the bandwidth of a network has improved substantially. This improvement in bandwidth has come from two sources – faster links and better topologies. In the same period of time, the latency of the network has not improved in this manner. In the time bandwidth doubles, latency improves by no more than a factor of 1.2-1.4 (Patterson, 2004). Many business applications, such as remote database access and interactive transaction processing, contain software timers that do not adjust for long delays. Examples include many banking applications, airline reservation and scheduling applications. This puts a serious limitation on networks to be able to support real-time applications.

Propagation delay is the primary source of latency. The Velocity Factor (a ratio of the speed at which a wave passes through the medium, to the speed of light in vacuum) of fiber optic cables is about 70 percent; whereas for a copper cable, it varies from 40 to 80 percent depending on the construct (O3b Networks, 2008). This is simple physics and nothing can be done in this regard. However, other factors like serialization delay, routing and switching latencies, latency induced by protocols, and so on, can be minimized to some extent. Unfortunately, latency also depends upon the geographical area (or physical distance) between the communicating devices. As the networks cover more and more geographical area, latency is expected to rise further.

Even though bandwidth is growing substantially in comparison to latency, yet, in relation to computing power, it still lags behind. Nielsen's law is similar to the more established Moore's Law. Unfortunately, comparing the two laws shows that bandwidth grows slower than computing power. Annual growth of processing power is about 60 percent; while as that of internet bandwidth is about 50 percent. Again, this puts networks under extreme pressure to cope up with the growing demands.

4.3 Trend in data analysis (in horizontal scaling)

The current trend in Big Data analysis is to use a network of servers connected together by high speed media. There are two factors behind this trend: low-cost infrastructure; and unlimited scaling. However, this implies that in order to expect timely results, the network must be able to support high bandwidth. In the early 1980s, 10 Mbps Ethernet (a system for connecting a number of computer systems to form a local area network) was very fast. By 1990s, another technology for networking, namely, Fiber Distributed Data Interface (FDDI) emerged which was was ten times faster than regular Ethernet (Abeysundara and Kamal, 1991). The next important iteration of Ethernet technology, namely, Gigabit Ethernet, was able to support a bandwidth of 1 Gbps, and sooner 10 Gbps Gigabit Ethernet came into existence (Kung, 1992). Without doubt, 100 Gbps was the next big leap, but transmitting at 100 Gbps over fiber has numerous challenges. Nevertheless, 40 and 100 Gbps Gigabit Ethernet standards exist by making use of parallel datastreams, wherein each stream uses a slightly different wavelength laser light. Under-sea cables already transport multi-terabit aggregate bandwidths over a single fiber using Dense Wavelength Division Multiplexing; so it seems development of Terabit Ethernet is inevitable (Van and Harmen, 1994).

On the other hand, in the recent past, in order to provide overall system bandwidth, and performance, and to be able to scale thousands of nodes, point-to-point switching fabric network solutions have been developed. There are two main advantages of switch fabric infrastructure compared to other architectures (Juniper Networks, 2012): first, there is a single hop between any two nodes which significantly reduces latency between them; and second, it provides a seamless high-performance interconnect with the easiest switching management.

Currently, Gigabit networks and fabric networks provide high performance and scalable solution for Big Data analysis in peer-to-peer networks like Hadoop and Spark.

4.4 Challenges in data analysis (in horizontal scaling)

Network considerations for Big Data analysis are not limited to bandwidth and throughput only. Availabilty is another important factor. In a system of interconnected servers, if the network is not available (due to hardware or software failure, or human error, or maintenance windows, and so on), the result is a collection of isolated compute resources and data sets which are rendered useless. To completely avoid downtime, networks need to be resilient to failures. Resilience in networks is determined by path diversity (having more than one way to get between resources) and fail over (being able to identify issues quickly and fail over to other paths). As there are many varied reasons for downtime, it is difficult to design a resilient and perfectly available network.

Congestion is another design consideration for Big Data networks which can lead to more than queuing delays and dropped packets. Congestion can trigger retransmissions, which can cripple already heavily loaded networks. Accordingly, networks need to be architected to mitigate congestion wherever possible. As with the design criteria for availability, mitigating congestion requires networks with high path diversity, which allows the network to fan traffic out across a large number of paths between resources.

In Big Data analysis, when compute time is measured in seconds or minutes, network latency of the order of nanoseconds is affordable. However, when the application is executed in parallel acrosss the servers, the communication between them needs to be highly synchronous, failing which overall application failure may be expected. Therefore, it is important that networks provide not just efficient but also consistent performance across both space and time.

Network partitioning is another crucial design parameter. By segregating the Big Data traffic from residual network traffic, the bursty demands from applications cannot impact other mission-critical workloads. Doing this requires networks to keep workloads logically separate in some cases and physically separate in others. Furthermore, depending on the application, the requirements may vary: Some might be particularly bandwidth-heavy, while as others might be latency-sensitive. Ultimately, a network that supports multiple applications, and multiple tenants, must be able to distinguish among their workloads and treat each appropriately.

5. Storage technology

Currently digital data storage is at the prime focus of every organization, like never before. Big Data and storage are being pronounced together in same breath. The fact is that due to high volume, velocity and variety of Big Data, storage subsystems are under great pressure to be able to store and retrieve huge volumes of data with reasonable performance. This section discusses the current technology trends in storage subsystems to support storage and processing of Big Data while highlighting the technological gaps between the trends and the requirements.

5.1 Trend in data storage

Since the inception of magnetic disk drive, there have been continuous design improvements in disk drives to support large capacity and high performance. The storage capacity of a magnetic disk drive is directly proportional to its areal density which is computed as the product of two other density measures – track density and linear density (track density measures how tightly the concentric tracks on the disk are packed; while as linear density measures how tightly the bits are packed within the length of a track). This density progress has been the result of a series of laboratory investigations of new head designs and technologies (Grochowski, 1998). The evolution of head design involves a reduction in sensor length (which increases track density), and a reduction in film thickness (which determines linear density). In 2003 (Morris and Truskowski, 2003) observed that since 1980 the areal density of magnetic disk drive has increased by seven orders of magnitude, while as the cost has declined by five orders. In the same time, (Grochowski and Halem, 2003) argued that with a Compound Growth Rate (CGR) of 60 to 100 percent per year in areal density of disk drive, the expected price-declines would average 37 to 50 percent per year. In 2009 (Kryder and Kim, 2009) argued that if the areal density of magnetic disks would continue to progress with the then-current CGR of 40 percent, a 14 TB drive will cost about $40 in year 2020. Figure 3 shows that annually the cost of magnetic disk drive is reducing by half, while as its capacity is doubling.

Every nation nowadays maintains mountains of digital data gathered from surveillance cameras, satellite images, logs maintained by ISPs and tens of terabytes of historical call data. A study by the University of California at Berkeley estimated that the amount of original digital content creation each year is about 1.6 million terabytes (Seagate, 2001). However, a mathematical model developed in a recent survey showed that in future the rate of production of digital data will stabilize at 13.2 ZB per year (Makarenko, 2011). Another research has estimated that the growth in amount of digital information created annually from 2009 to 2020 is about 44 times (Moore, 2009). The current trend in storage technology shows that the capacity of magnetic disk drives is increasing, while as storage cost per unit bit is decreasing, thereby allowing the organizations to store large amount of digital data at affordable price.

However, magnetic disk drives are mechanical at heart, and thus the disks' rotational speed has slightly improved over the last decade. This implies that disks will remain significantly faster for sequential accesses than for random accesses, and the gap will only grow. This can severely limit the performance that hard disk-based systems are able to offer to workloads with significant random access component (or lack of locality). Optimizing data access is a popular way to improve the

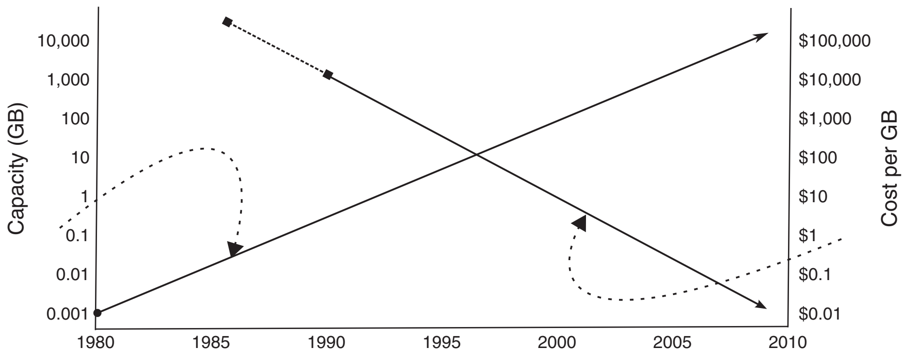

Figure 3.
Growth in capacity and reduction in cost/bit of magnetic storage

performance of data-intensive computing (Ishii and Fernandes de Mello, 2012; Ishii and de Mello, 2011; Ishii and de Mello, 2009). These techniques include data replication, access parallelism, migration and distribution.

Using storage Redundant Array of Inexpensive Drives (RAID) to improve performance and fault tolerance has proven promising in the recent past (Chen *et al.*, 1994). RAID employs data-mirroring in which same data were stored across two hard drives so that it can independently process two disk access requests at the same time. The downside to mirroring is that the capacity is only half of the total capacity of all hard drives, so it is expensive.

Direct-attached storage (DAS), network-attached storage (NAS) and storage area network (SAN) are the enterprise storage architectures that are commonly used (Leong, 2009). DAS refers to block-oriented storage which is directly attached to a server. SAN provides block-oriented storage that resides across a network. NAS also provides access to storage residing across a network, but accessible via a higher level interface such as files or objects. However, all these architectures lack the essential requirements for the applications on highly scalable computing clusters (Chen and Zhang, 2014).

On the other side of the spectrum, the continued improvement in the cost and performance of flash-based storage has made solid state disks (SSDs) a viable technology. Flash Translation Layer, which is a technique that prolongs SSD life by ensuring wear-leveling and provides high writing speed, is a key to SSD performance. As an example, it is found that for certain random write-dominated workloads, the overheads of garbage collector and wear-leveling can sometimes make SSDs slower than HDDs (Lee and Moon, 2007). For sequential accesses, HDDs can easily outperform SSDs. Nevertheless, SSDs hold a lot of potential for higher and more predictable performance than HDDs. Although SSDs may be useful as stand-alone secondary storage for very high throughput and low-latency applications, they are generally expected to remain in the supporting role for hard disks in the foreseeable future. Many papers have explored SSD as an intermediate layer in the storage hierarchy between main memory and HDD-based secondary storage (aDam LeVenthaL, 2008; Narayanan *et al.*, 2009), and as a storage device working in tandem with HDD (Fisher *et al.*, 2012; Bhat and Quadri, 2014).

Furthermore, there are a host of other Non-Volatile RAM (NVRAM) technologies under development that may significantly alter the storage landscape in the near future. Some of the more prominent NVRAM technologies include Magnetic RAM (MRAM), Phase-Change Memory (PRAM or PCM) and Ferroelectric RAM (FeRAM) (Muller *et al.*, 2004; Kohlstedt *et al.*, 2005). These NVRAM technologies offer several advantages over the rotating magnetic media: lower and more predictable access latencies for random requests, smaller form factors, lower power consumption, lack of noise, and higher robustness to vibrations and temperature.

5.2 Challenges in data storage

More commonly known as Kryder's law, it was then predicted that the areal storage density of a magnetic disk drive doubles approximately after every 18 months, like the doubling of transistor count after every 18 months in Moore's law. However, the validity of the Kryder's law's projection of 2009 was questioned halfway into the forecast period. In past, more such predictions had been made; the less familiar but noteworthy being Storage law (Porter, 2005). The study observed that the capacity of magnetic storage then doubled after every nine months for at least a decade. It was twice the rate predicted by Moore's law for growth of computing power. Other studies concluded that our ability to capture and store data far outpaces our ability to process and exploit it (Fayyad and Uthurusamy, 2002). Unfortunately, there is no golden yardstick in magnetic storage technology which can perfectly (or near perfectly) predict the growth of magnetic storage capacity. Furthermore, as Big Data applications deal with large volumes of information, employing replication for resilience or mirroring for performance (as in case of RAID) means that the Big Data just got bigger.

Generally, magnetic disk drives were used to store persistent data (Kasavajhala, 2011) and due to their much slower random I/O performance than sequential I/O performance, the applications were restructured to work around this limitation. Now, magnetic disks are being replaced by SSDs and other technologies such as PCM (Agrawal *et al.*, 2011). These storage devices do not have the same large spread in performance between the sequential and random I/O. Therefore, the earlier performance patch may actually add overhead, and hence, may deteriorate the performance of non-magnetic storage drives.

In NVRAM-based storage, it is necessary to re-examine the traditional distinction between main memory and secondary storage access. When a thread stalls for disk access, the OS takes control and switches the thread to another process since the latency of I/O completion is very large compared with the cost of a context switch. However, such a switch does not make sense for a memory access. With extremely fast solid state storage like NVRAM, such a distinction may no longer hold and an adaptive context switch mechanism may be required. Furthermore, a fast storage access exposes the high overhead of the traditional file-system layer, and it is necessary to re-examine traditional file-access model to make it substantially leaner. (Lee *et al.*, 2009) and (Zhou *et al.*, 2009) examine multi-level memory system using NVRAM to assess the performance benefits of this approach.

6. Discussion and conclusion

Big Data offers wide range of opportunities to academics, industry, businesses, research and so on. It relies on the technology for its growth and analysis, in order to deliver the value it promises. One promise of Big Data analysis is knowledge creation and the economy built on it. For a sustainable knowledge economy to be an outcome of Big Data, technology must play its part effectively both at Big Data source and sink. At Big Data source, technology must keep on refreshing Big Data along all its three major axis, i.e., volume, variety and velocity (Laney, 2001). Otherwise, with time, Big Data would become stagnant and knowledge, so derived, obsolete. Similarly, technology must facilitate correct, timely and diverse analysis at Big Data sink, failing which agile, stable and strong knowledge economy could not be expected. However, as technology has evolved keeping in view many factors but Big Data, technological gaps creep in at various stages of Big Data.

The study aims to identify the challenges posed by Big Data (at various stages of Big Data analysis) to current trends in three main aspects of hardware technology: computation; networking and storage technology. The motive is to bridge the gap between theory and practice, contribute to knowledge base, identify potential areas of research and stage-specific effective-technology, and so on. The study employs a methodical and critical review of relevant literature as a tool to accomplish the task.

The results suggest that current trends in hardware technology have not evolved with the motivation to facilitate generation, transportation and analysis of Big Data, but it considerably and significantly supports Big Data analysis at all stages. However, there are certain areas of the technology that need attention with respect to Big Data.

Computation technology would further widen the domain of applications of digital technologies by providing low cost, small size, high-performance and power efficient CPUs. This implies that it would allow researchers to envision new and novel applications besides facilitating practitioners to realize existing theoretical-applications. Furthermore, it would allow digital technologies to penetrate further deep into every aspect of our life: all facilitating growth and proliferation of digital data. As a result, academics and industry would have heavier, richer and aggressive Big Data at their disposal to retrieve more knowledge, new kinds of knowledge and other values from Big Data. At the same time, the technology provides a wide variety of application-specific computing configurations for Big Data analysis which includes GPUs, HPC and FPGA for real-time analysis, and peer-to-peer networks for scalability. This suggests that researchers and practitioners could perform diverse analysis of Big Data to retrieve any specific value from it. Besides, existence of low-cost peer-to-peer networks allows low-budget enterprises to exploit the benefits of Big Data. Nevertheless, Big Data still challenges computation technology. Big Data sources are not intelligent enough to avoid generation of erroneous data, and fails to ensure generation of metadata for the correct data generated. This leads to Big Data pollution, and makes Big Data analysis more computational intensive. As a result, Big Data analysis may be misleading, not performed in timely manner and may yield inappropriate results. However, augmenting Big Data sources with this capability would make the analysis more accurate, timely and effective. This provides researchers with an opportunity to explore ideas to filter erroneous data, and practitioners with the knowledge of existence of possible inappropriate results. Moreover, the technology fails to provide a single computing-configuration for both real-time and scalable analysis. This forces an enterprise to compromise on performing diverse analysis of Big Data, and restricts practitioners to a specific set of Big Data analysis; all depending on the investments made. Again, it provides researchers an opportunity to explore solutions, at both hardware and software level.

Networking technology continues to cover more and more geographical area while consistently improving bandwidth and reducing cost. This implies that for Big Data acquisition, the digital-geographical boundaries would diminish, thereby making Big Data richer and heavier. Therefore, researchers and practitioners can perform diverse Big Data analysis over a data set which represents whole world, as well performing regional analysis. On the other hand, Gigabit and fabric networks provide low-cost high-performance scalable solution for Big Data analysis in peer-to-peer networks, like Hadoop and Spark. This means enterprises with low-budget can also reap benefits out of Big Data. However, networking technology too has its limitations. The growing void between bandwidth and latency in networking technology hinders the capability of networks to support real-time applications. The growing gap between the CPU processing power and network bandwidth further exaggerates it. This suggests that researchers and practitioners cannot exploit Big Data to its full potential. Worse, when the latency in networks crosses some threshold value, the analysis is limited to non-real-time applications. At the other end of the spectrum, in peer-to-peer networks, features like availability, congestion, synchronization and network partitioning are more dominant, and are unfortunately missed in Gigabit and fabric networks. This implies that practitioners and researchers would have to consider these parameters while choosing a specific analysis of Big Data. Nevertheless, it provides an opportunity to researchers to explore new theories and existing practices, both at hardware and software level.

Finally, storage technology has evolved to a stage where it could cater to the demand of low-cost, huge-storage and high-performance digital storage for Big Data. Huge capacity magnetic disk drives come in multiple configurations like DAS, RAID, NAS and SAN. On the other hand, for performance purpose, SSDs are available. However, practitioners can use SSDs and NVRAMs in tandem with magnetic drives to deliver high performance with cost-effective huge-capacity. This implies that practitioners could store huge volumes of data at low-cost and could retrieve the same with the needed speed. This is useful in Big Data analysis as the complete data set would be available for analysis with affordable latency. However, storage technology has its shortcomings. In storage technology, due to lack of well-established rule for predicting the growth of storage capacity of magnetic drives, the model for magnetic capacity growth needs rethinking. This suggests that researchers should explore new ideas to accommodate aggressively growing Big Data before the storage costs supersede the value derived out of Big Data stored. The support for resilience and performance in magnetic disks (e.g. mirroring in RAID) demands at least double the storage space required by Big Data. This further worsens the problem. Also, using SSDs and NVRAMs, alone, or in tandem with magnetic drives, brings up a whole new set of problems influencing every piece of hardware and software. This demands new investments in software and hardware. Nevertheless, this provides an opportunity to researchers to explore new ideas to find out low-cost overheads for integration of SSDs, NVRAMs and magnetic disks.

This study provides a comprehensive and concise technical insight into the current hardware technology trends employed in Big Data analysis. This knowledge would be of great help to industry, organizations and enterprises to understand what to expect and what not to expect from current technology trends in solving Big Data problems. And, therefore, would arm practitioners with the knowledge to choose meticulously a specific hardware for a specific analysis. At the same time, the study provides an opportunity to researchers to address a wide variety of challenges that need to be addressed in current trends of computation, networking and storage technology employed by Big Data. It is hoped that the article will provide researchers many interesting avenues to explore for meeting the end user expectation of Big Data analysis and capabilities of hardware technology to deliver it.

References

Abeysundara, B.W. and Kamal, A.E. (1991), "High-speed local area networks and their performance: a survey", *ACM Computing Surveys*, Vol. 23 No. 2, pp. 221-264.

aDam LeVenthaL, B.Y. (2008), "Flash storage memory", *Communications of the ACM*, Vol. 51 No. 7, pp. 47-51.

Agrawal, D., Das, S. and El Abbadi, A. (2011), "Big data and cloud computing: current state and future opportunities", *Proceedings of the 14th International Conference on Extending Database Technology, ACM, Uppsala, March 21-24*, pp. 530-533.

Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M. and Widom, J. (2011), "Challenges and opportunities with Big Data", technical report, Purdue University, available at: http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1000&context=cctech (accessed October 3, 2014).

Bayo-Moriones, A., Billón, M. and Lera-López, F. (2013), "Perceived performance effects of ICT in manufacturing SMEs", *Industrial Management & Data Systems*, Vol. 113 No. 1, pp. 117-135.

Begona Lloria, M. and Peris-Ortiz, M. (2014), "Knowledge creation. The ongoing search for strategic renewal", *Industrial Management & Data Systems*, Vol. 114 No. 7, pp. 1022-1035.

Bhat, W.A. and Quadri, S.M.K. (2014), "Performance augmentation of a FAT filesystem by a hybrid storage system", *Proceedings of Advanced Computing, Networking and Informatics-Volume 2, Springer International Publishing*, pp. 489-498.

Borkar, S. (2007), "Thousand core chips: a technology perspective", *Proceedings of the 44th annual Design Automation Conference, ACM*, pp. 746-749.

Brown, S.D., Francis, R.J., Rose, J. and Vranesic, Z. (2012), *Field Programmable Gate Arrays*, Springer Science & Business Media, New York, NY.

Buyya, R. (1999), *High Performance Cluster Computing: Architecture and Systems, Volume I*, Prentice Hall, Upper SaddleRiver, NJ.

Chen, C.P. and Zhang, C.-Y. (2014), "Data-intensive applications, challenges, techniques and technologies: a survey on Big Data", *Information Sciences*, Vol. 27 No. 5, pp. 314-347.

Chen, P.M., Lee, E.K., Gibson, G.A., Katz, R.H. and Patterson, D.A. (1994), "RAID: high-performance, reliable secondary storage", *ACM Computing Surveys (CSUR)*, Vol. 26 No. 2, pp. 145-185.

Durairajan, S., Prado, M.I., Rahimi, N. and Jahromi, S.R. (2013), "Forecasting microprocessor technology in the multicore era using TFDEA", *PICMET'13: Proceedings of Technology Management in the IT-Driven Services, IEEE*, pp. 2108-2115.

Fayyad, U. and Uthurusamy, R. (2002), "Evolving data into mining solutions for insights", *Communications of the ACM*, Vol. 45 No. 8, pp. 28-31.

Fisher, N., He, Z. and McCarthy, M. (2012), "A hybrid filesystem for hard disk drives in tandem with flash memory", *Computing*, Vol. 94 No. 1, pp. 21-68.

Francis, R.J., Rose, J. and Vranesic, Z. (1992), *Field Programmable Gate Arrays*, Springer Science & Business Media, New York, NY.

Franklin, M.J. (2013), "Making sense of big data with the Berkeley data analytics stack", *25th International Conference on Scientific and Statistical Database Management, ACM*, pp. 2-3.

Grochowski, E. (1998), "Emerging trends in data storage on magnetic hard disk drives", technical report, Datatech, ICG Publishing, available at: https://classes.soe.ucsc.edu/cmps129/Winter03/papers/grochowski-trends.pdf (accessed February 15).

Grochowski, E. and Halem, R.D. (2003), "Technological impact of magnetic hard disk drives on storage systems", *IBM Systems Journal*, Vol. 42 No. 2, pp. 338-346.

Haynes, P. and NGuyen, C.M.-H. (2014), "Rebalancing socioeconomic asymmetry in a data-driven economy", The Global Information Technology Report, available at: www3.weforum.org/docs/GITR/2014/GITR_Chapter1.6_2014.pdf (accessed December 10, 2014).

Hofstee, H.P. (2005), "Power efficient processor architecture and the cell processor", *HPCA-11: Proceedings of 11th International Symposium on High-Performance Computer Architecture, IEEE*, pp. 258-262.

Internetworldstats (2015), available at: www.internetworldstats.com/stats.htm (accessed August 21).

Isci, C., Buyuktosunoglu, A., Cher, C.Y., Bose, P. and Martonosi, M. (2006), "An analysis of efficient multi-core global power management policies: maximizing performance for a given power budget", *Proceedings of the 39th annual IEEE/ACM international symposium on microarchitecture, IEEE Computer Society*, pp. 347-358.

Ishii, R.P. and de Mello, R.F. (2009), "A history-based heuristic to optimize data access in distributed environments", *Proceedings of the 21st IASTED International Conference, ACTA Press*, p. 198.

Ishii, R.P. and de Mello, R.F. (2011), "An adaptive and historical approach to optimize data access in grid computing environments", *INFOCOMP Journal of Computer Science*, Vol. 10 No. 2, pp. 26-43.

Ishii, R.P. and Fernandes de Mello, R. (2012), "An online data access prediction and optimization approach for distributed systems", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 23 No. 6, pp. 1017-1029.

Jianzhong, L. and Xianmin, L. (2013), "An important aspect of big data: data usability", *Journal of Computer Research and Development*, Vol. 2013 No. 6, pp. 1147-1162.

Juniper Networks (2012), "Introduction to Big Data: Infrastructure and Networking Considerations", Juniper White Paper, available at: www.juniper.net/us/en/local/pdf/whitepapers/2000488-en.pdf (accessed November 7, 2014).

Kasavajhala, V. (2011), "Solid state drive vs hard disk drive price and performance study", Dell Tech. White Paper, available at: www.itc.vn/attachments/article/288/ssd_vs_hdd_price_and_performance_study.pdf (accessed February 15).

Khan, F. and Pi, J. (2011), "Millimeter-wave mobile broadband: unleashing 3-300 GHz spectrum", *Presented at IEEE Wireless Communication and Networking Conference, Princeton, NJ, March*.

Kmieciak, R., Michna, A. and Meczynska, A. (2012), "Innovativeness, empowerment and IT capability: evidence from SMEs", *Industrial Management & Data Systems*, Vol. 112 No. 5, pp. 707-728.

Kohlstedt, H., Mustafa, Y., Gerber, A., Petraru, A., Fitsilis, M., Meyer, R. and Waser, R. (2005), "Current status and challenges of ferroelectric memory devices", *Microelectronic Engineering*, Vol. 80 No. 1, pp. 296-304.

Kryder, M.H. and Kim, C.S. (2009), "After hard drives – what comes next?", *IEEE Transactions on Magnetics*, Vol. 45 No. 10, pp. 3406-3413.

Kshetri, N. (2014), "Big Data? Its impact on privacy, security and consumer welfare", *Telecommunications Policy*, Vol. 38 No. 11, pp. 1134-1145.

Kung, H.T. (1992), "Gigabit local area networks: a systems perspective", *IEEE Communications Magazine*, Vol. 30 No. 4, pp. 79-89.

Laney, D. (2001), "3D data management: controlling data volume, velocity and variety", *META Group Research Note*, 70pp.

Lee, B.C., Ipek, E., Mutlu, O. and Burger, D. (2009), "Architecting phase change memory as a scalable dram alternative", *ACM SIGARCH Computer Architecture News*, Vol. 37 No. 3, pp. 2-13.

Lee, S.W. and Moon, B. (2007), "Design of flash-based DBMS: an in-page logging approach", *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 55-66.

Leong, D. (2009), "A new revolution in enterprise storage architecture", *IEEE Potentials*, Vol. 28 No. 6, pp. 32-33.

Little, G. (2012), "Managing the data deluge", *The Journal of Academic Librarianship*, Vol. 38 No. 5, pp. 263-264.

Lynch, C. (2008), "Big data: how do your data grow?", *Nature*, Vol. 455 No. 7209, pp. 28-29.

Madden, S. (2012), "From databases to big data", *IEEE Internet Computing*, Vol. 16 No. 3, pp. 4-6.

Makarenko, A.V. (2011), "Phenomenological model for growth of volumes of digital data", *arXiv preprint arXiv: 1102.5500*, available at: http://arxiv.org/ftp/arxiv/papers/1102/1102.5500.pdf (accessed October 22, 2014).

Milojicic, D., Kalogeraki, V., Lukose, R., Nagaraja, K., Pruyne, J., Richard, B. and Xu, Z. (2002), "Peer-to-peer computing", Technical Report No. HPL-2002-57, HP Labs, available at: http://141.223.82.74/p2p/papers/HPL-2002-57.pdf (accessed December 25, 2014).

Moore, F. (2009), "Storage facts, figures, best practices, and estimates", *Horison Information Strategies*, available at: www.slideshare.net/datacenters/storage-facts-and-figures-msword (accessed December 15, 2014).

Moore, G.E. (2006), "Cramming more components onto integrated circuits", *Solid-State Circuits Society Newsletter*, Vol. 11 No. 5, pp. 33-35.

Morris, R.J. and Truskowski, B.J. (2003), "The evolution of storage systems", *IBM systems Journal*, Vol. 42 No. 2, pp. 205-217.

Mukherjee, D. and Krishnamurthy, K. (2012), "Maximizing the returns from big data", *Cognizanti*, Vol. 5 No. 1, pp. 1-10.

Muller, G., Happ, T., Kund, M., Lee, G.Y., Nagel, N. and Sezi, R. (2004), "Status and outlook of emerging nonvolatile memory technologies", *IEEE International Electron Devices Meetng, IEEE*, pp. 567-570.

Narayanan, D., Thereska, E., Donnelly, A., Elnikety, S. and Rowstron, A. (2009), "Migrating server storage to SSDs: analysis of tradeoffs", *Proceedings of the 4th ACM European Conference on Computer Systems, ACM*, pp. 145-158.

Nielsen, J. (1998), "Nielsen's law of internet bandwidth", available at: www.useit.com/alertbox/980405.html (accessed December 10, 2014).

O3b Networks (2008), "What is network latency and why does it matter?", available at: www.o3bnetworks.com/wp-content/uploads/2015/02/white-paper_latency-matters.pdf (accessed December 20, 2014).

Ong, C.S. and Chen, P.Y. (2014), "The effects of IT: from performance to value", *Industrial Management & Data Systems*, Vol. 114 No. 1, pp. 70-85.

Owens, J.D., Luebke, D., Govindaraju, N., Harris, M., Krüger, J., Lefohn, A.E. and Purcell, T.J. (2007), "A survey of general-purpose computation on graphics hardware", *Computer graphics forum*, Vol. 26 No. 1, pp. 80-113.

Patterson, D.A. (2004), "Latency lags bandwith", *Communications of the ACM*, Vol. 47 No. 10, pp. 71-75.

Pollack, F.J. (1999), "New microarchitecture challenges in the coming generations of CMOS process technologies", *Proceedings of the 32nd annual ACM/IEEE international symposium on Microarchitecture, IEEE Computer Society*, p. 2.

Porter, J. (2005), "Disk/trend reports 1977-1999", presented at the THIC Meeting at the Sony Auditorium, San Jose, CA, April 19-20, available at: www.thic.org/pdf/April05/disktrend.jporter.050419.pdf (accessed February 22).

Rao, B. and Angelov, B. (2005), "Bandwidth intensive applications: demand trends, usage forecasts, and comparative costs", NSF-ITR technical report on fast file transfers across optical circuit-switched networks, available at: http://faculty.poly.edu/~brao/2005.BandwidthApps.NSF.pdf (accessed January 13).

Santosh, P.G. and Bhandare, K. (2014), "Next generation infrastructure for Big Data-a challenge", available at: www.aisectuniversity.ac.in/Anusandhan6/17_Praveen_Goyal_Research_Paper_15.05.14.pdf (accessed November 24, 2014).

Seagate (2001), "Disc drive capacity and performance", Seagate White Paper, available at: www.seagate.com/docs/pdf/whitepaper/disc_capacity_performance.pdf (accessed January 18).

Seiler, L., Carmean, D., Sprangle, E., Forsyth, T., Abrash, M., Dubey, P. and Hanrahan, P. (2008), "Larrabee: a many-core x86 architecture for visual computing", *ACM Transactions on Graphics*, Vol. 27 No. 3, 15pp.

Shvachko, K., Kuang, H., Radia, S. and Chansler, R. (2010), "The hadoop distributed file system", *IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), IEEE*, pp. 1-10.

Steinmetz, R. and Wehrle, K. (2005), *Peer-to-Peer Systems and Applications*, Lecture Notes in Computer Science (LNCS), Springer Science & Business Media, Springer Berlin Heidelberg.

Stuart, J.A. and Owens, J.D. (2011), "Multi-GPU MapReduce on GPU clusters", *Proceedings of 2011 IEEE International Parallel & Distributed Processing Symposium (IPDPS), IEEE*, pp. 1068-1079.

Tankard, C. (2012), "Big data security", *Netwrok Security*, Vol. 2012 No. 7, pp. 5-8.

Thomas, D. and Moorby, P. (2002), *The Verilog Hardware Description Language, Vol. 2*, Springer Science & Business Media, Springer.

Van, A. and Harmen, R. (1994), "Media access techniques: the evolution towards terabit/s LANs and MANs", *Computer Networks and ISDN Systems*, Vol. 26 No. 6, pp. 603-656.

Vesset, D., Woo, B., Morris, H.D., Villars, R.L., Little, G., Bozman, J.S. and Yezhkova, N. (2012), "Worldwide big data technology and services 2012-2015 forecast", Report No. 233485, IDC.

Zhou, P., Zhao, B., Yang, J. and Zhang, Y. (2009), "A durable and energy efficient main memory using phase change memory technology", *ACM SIGARCH Computer Architecture News*, Vol. 37 No. 3, pp. 14-23.

About the authors

Dr Wasim Ahmad Bhat is an Assistant Professor of Computer Science at the University of Kashmir. He received his PhD in computer science from the University of Kashmir. His research interests include operating systems, specifically file systems and storage subsystems. Dr Wasim Ahmad Bhat is the corresponding author and can be contacted at: wab.cs@uok.edu.in

Professor S.M.K. Quadri is a Professor of Computer Science at the University of Kashmir. He received his PhD in computer science from the University of Kashmir. His research interests include software reliability, testing and other design considerations.