



# Price Versus Performance of Big Data Analysis for Cloud Based Internet of Things Networks

Khalim Amjad Meerja<sup>1</sup> · Praveen V. Naidu<sup>1</sup> · Sri Rama Krishna Kalva<sup>1</sup>

Published online: 23 May 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

Many devices share their data with the online world to derive global knowledge and information that have high business value. Trillions of smart devices are connected together over the Internet, which are known as Internet-of-Things (IoT). These devices generate enormous data on daily basis, in orders of Exabytes, which is called Big Data. Since cloud services are used to handle the Big Data generated from these IoT devices, new architectures for handling smart devices are designed through cloud enabled IoT networks. In this paper, we discuss in detail the issues of handling Big Data from an operational perspective in this new cloud based IoT network architecture. We tackle the incurred price and overall efficiency for storing and analyzing data for these networks on periodical basis. We propose an optimization model that address the price versus performance while carrying out Big Data analysis in these cloud based IoT networks.

**Keywords** Internet of things (IoT) · Big data · Mobile cloud computing (MCC) · Cloud radio access network (C-RAN) · Software defined networks (SDN) · Network function virtualization (NFV)

## 1 Introduction

Internet of Things (IoT) are heterogeneous, fixed, mobile, miniature, small, large, active, and passive radio frequency based communicating data sources. They are embedded in trillions of devices present in the surrounding environment, which will generate Big Data [1]. For instance, it is now a reality that most of the merchandize products are equipped with radio frequency identification (RFID) tags. They send information to a central tag reader, which is a wireless base station that collects information on the goods such as quantity, price, date of expiry, etc. Based on the collected information, the supervisor in the store can order new stock, place sale on items, re-arrange items in the store for better public attraction, and take any other actions that will enhance sales and profitability.

Another eye catching reality is the new generation of smart homes, which can be managed through an appropriate remote controller or simply using an application on the mobile phone. This is made possible with the help of smart sensors embedded in home appliances. These sensors are capable of regulating the home functionality such as adjusting the room temperature, controlling home ventilation and lighting, placing grocery orders to the nearest or selected store, etc. Another impending application of smart IP devices installed in homes is the remote health care to save human lives. The devices adorned by humans will monitor their health and periodically send critical health information to the nearest concerning hospital or directly to a doctor's device depending on the type of emergency.

There are many applications of using smart sensors. A survey of some of these applications can be obtained from [2–4]. Many sensor devices generate data while monitoring a specific activity. This data is used to create new applications that save lives, improve quality of living, enhance security, provide impetus for future lifestyles and habits, and predict new demands. This means an explosive growth in the number of sensor devices, possibly in trillions, that are interconnected to each other and ultimately to the Internet backbone. More importantly, is the huge unstructured data that these dissimilar IoT devices generate on daily basis in order of Exabyte.

---

This work is partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC-STPGP 447230).

---

✉ Khalim Amjad Meerja  
kmeerja@gmail.com

<sup>1</sup> Department of Electronics and Communications,  
V R Siddhartha Engineering College, Vijayawada,  
Andhra Pradesh, India

New challenges emerge while handling this data explosion, termed as Big Data, produced by IoT devices around the globe. New scalable and efficient architectures are needed and should cater according to the public needs such as social, industrial, business, scientific, and security [5–10]. Big Data processing and analysis is challenging because they need innovative solutions to process the huge amount of data, which is in order of Exabyte, according to the quality of service (QoS) requirements of the end applications. Sometimes the processing should be done in real-time, which needs appropriate software processing and data management tools [11]. Traditional relational database management system (RDBMS) tools are not sufficient to handle such huge data, even when the data is adequately structured. Therefore, this work does not delve in the actual data processing and data mining techniques. However, it considers the network requirements for achieving Big Data analysis efficiently and cost effectively [12].

Network infrastructure has new challenges to face in order to maintain the throughput and delay constraints required by end applications analyzing Big Data. The transmission control protocol (TCP), user datagram protocol (UDP), and Internet protocol (IP) currently have many limitations when handling streaming and elastic data traffic according to required QoS constraints. Therefore, new scalable network architectures are needed to overcome these limitations [13, 14]. These new architectures should address massive storage requirements that are swift in performing reading and writing operations. The current storage technologies are extremely slow compared to the available processor speeds [15]. This is a main problem when huge volumes of data has to be fetched and processed for real-time analysis. This limitation has to be addressed through some means of parallel storage and simultaneous retrieval from several storage devices.

Another important concern is the intermittent network failures that occur during processing and analysis of Big Data [16]. It will be extremely time consuming to restart processes of such large magnitudes, every time a network failure occurs. Efficient and seamless recovery mechanisms are essential while processing Big Data. The networks should be highly adaptable and reconfigurable due to the adhoc nature of the location of IoT devices, particularly in scenarios where mobile devices constitute majority of the IoT devices [17].

As standardization efforts are already taking place to address IoT infrastructure communication needs [18], there is a dire requirement for handling the enormous data generated by the massive IoT infrastructure. This paper focuses on how to handle the data generated from massive IoT infrastructure that has high business value. In particular, we investigate a cloud-based IoT network architecture to improve price versus performance while carrying out Big

Data analysis. We provide intuitive analysis through a practical example that helps understanding the necessity to take critical decisions while performing data analysis. Also, an optimization model is proposed to minimize the price while satisfying the performance constraints.

The rest of the paper is organized as follows. Section 2 presents various applications, which illustrate the high value of Big Data analysis. A cloud-based solution that addresses various issues is discussed Section 3. All the issues that are discussed above, including other challenges are discussed in this section. Then, Section 4 discusses the various aspects that must be considered while performing Big Data Analysis. To effectively handle Big Data analysis, an optimization model to minimize price versus performance is proposed in Section 5. Finally, Section 6 concludes the paper.

## 2 Big data analysis in IoT networks

In order to understand the need for Big Data analysis, we need to know from where the data is generated and how critical and important is the data for human beings. For instance, imagine a satellite providing raw data on the potential meteoroids that will hit the earth. We immediately understand the importance of the data because it is essential to track such meteoroids and take proper action. Another common day example is the data collected on surrounding climate all over the planet. The data is useful in predicting the climate changes, the changes in wind currents, and ultimately the weather conditions. Another important aspect is the health care sector where data is collected and processed to predict any epidemic disease outbreak in the country and take necessary precautions and prepare immunizations accordingly. All these are prime examples where data is collected through some means, processed, and then analyzed. This sort of data analysis is now done by business organizations to improve their product sales, services, and support [19].

IoT devices such as sensors, RFIDs, and smart phones are used to collect data to predict social behavior such as their habits, favorite choices, and lifestyle. These IoT devices are also used to provide security and manage chaotic conditions during disaster recovery and any other social disturbances. The data is also collected for scientific purposes to improve product performance, safety, and reliability. Most of these applications need the entire bulk of data for batch processing and then analyzing the entire processed collected data. In some situations, the data should be analyzed in real-time such as market stocks and emergency operations.

The IoT devices are limited in storage and processing capacities. They generate continuous smaller sized data. The data should be forwarded to appropriate storage locations. These locations either store data from many devices in

raw form or correlate and process the collected data before storing in the final form to remove any sort of redundancy. Such form of collective processed data saves storage space substantially. Still, the resultant data collected is huge, therefore, it is generally stored in cloud network. This historically collected data is then analyzed on a regular basis using appropriate data analyzing application to derive knowledge and information. The acquired knowledge and information are applied to the end users' business and scientific processes to achieve better and desirable results.

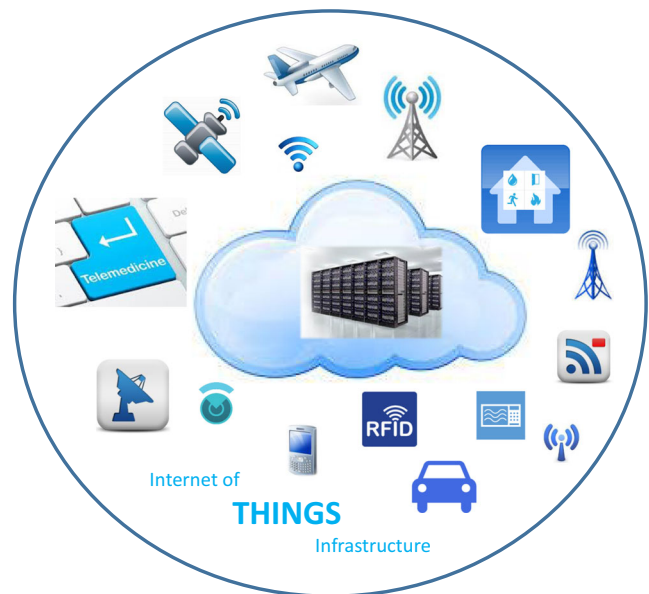
It is well known that the three important characteristics of the Big Data that is generated from these IoT devices are the sheer volume, variety, and velocity ("3Vs") [20]. These three characteristics must be understood and used for the benefit of the society and human life improvement. The applications have to be developed accordingly, looking into the direct consumer needs. The common public should benefit from the available Big Data and analysis tools. More importantly, the appropriate data must be available for access by the right people anywhere around the world. Technologies must be capable of promoting such goal into ultimate reality.

Some important wireless technologies that provide IoT devices for general public usage, which will produce Big Data are wireless multi-hop sensor/mesh networks, low power personal area networks (6LoWPAN), wireless local area networks (WLANs), 4G/5G/LTE-A and WiMAX cellular networks, machine-to-machine/device-to-device (M2M/D2D) networks, radio frequency identification (RFID) tags, near field communications (NFC), Bluetooth, and ZigBee [21–24]. The sample applications that are provisioned over these kind of IoT networks and their associated devices are shown in Fig. 1. Key applications are smart homes, telemedicine, smart tags for inventory tracking, e-commerce, social networks, health monitoring, terrestrial and satellite based monitoring, aviation control, etc.

The main limitations of many IoT devices are the limited battery life, limited storage, and less processing power. These disparate sources of data will not be able to provide global information and knowledge to large business organizations and scientific research communities. During data analysis, the need for global knowledge in these scenarios requires services of a powerful platform having high processing power and storage capacities. Therefore, IoT devices, particularly the wireless-based ones, are connected to cloud networks for back-end processing and data analysis [25].

### 3 Cloud based solution for IoT network infrastructure

A general overview of cloud support to IoT technologies is summarized in Table 1. The adoption of cloud solution



**Fig. 1** Cloud based IoT technologies and applications. There are many key applications to be mentioned. The figure shows applications such as smart home, telemedicine, inventory tracking, satellite communications, etc. Some of the applications have been summarized in Table 1

has been recommended for elastic provisioning of its services and handling the dynamic requirements of massive data storage and processing. The cloud network as a backbone for IoT infrastructure is ideal solution for processing and storing massive data, which is at least in the orders of Exabyte. In this line, a storage framework for Big Data generated from IoT infrastructure using cloud network is proposed in [26]. Similarly, a computational architecture for cloud based IoT infrastructure consisting of a manufacturing system is studied in [27]. The communication and quality of service aspects for such cloud based IoT infrastructure can be obtained from recent works [28, 29].

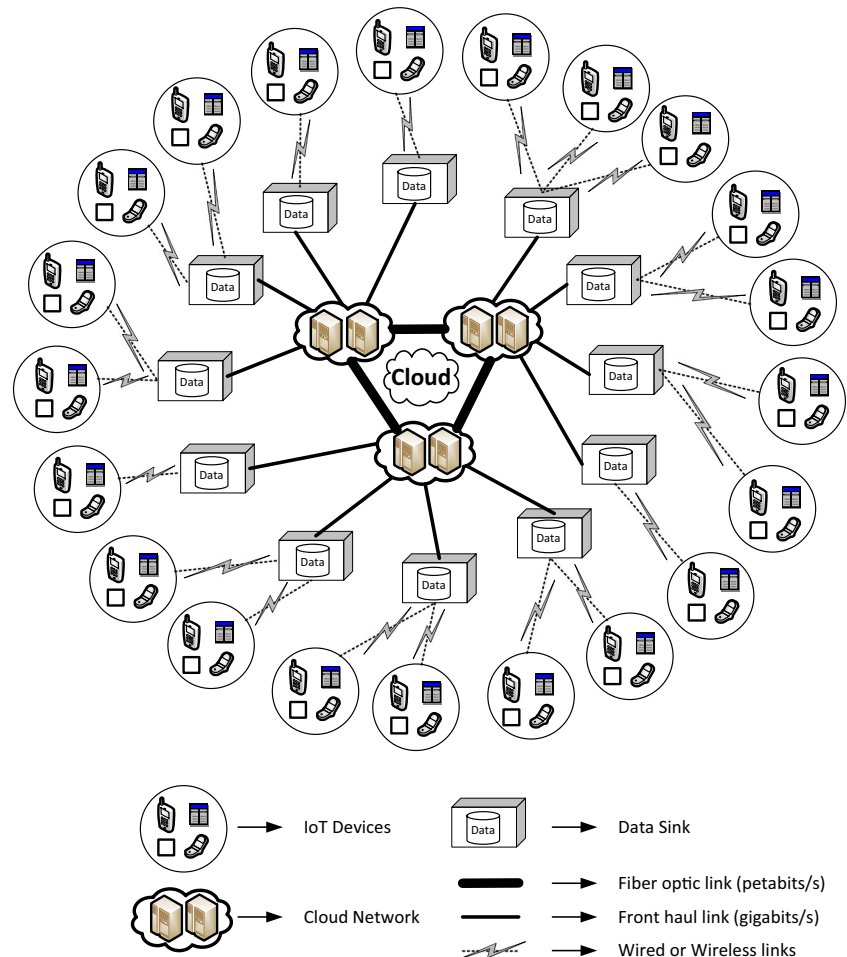
However, all these works have overlooked the impact of different policies of handling the massive data on the cost of performing data analysis while satisfying other functionality and green communication objectives. In this work, we attempt to address these objectives in a very specific aspect. Therefore, we adopt a generally applicable architecture for a cloud centered IoT infrastructure.

A scenario for current IoT infrastructure is shown in Fig. 2. IoT infrastructure consists of various kinds of fixed as well as mobile embedded devices, sensors, RFIDs, smart phones, and any other data sources with limited storage and computational power. They forward their data continuously to the nearest data sink through a wired link or over a wireless radio frequency channel. The data sink temporarily stores all the data that is received from its associated IoT devices. Then the sink

**Table 1** Key enabling technologies for the next generation wireless networks including 5G networks

IoT networks	Big data sources	Major applications	Cloud support
Wireless Sensor Networks (WSN)	<ol style="list-style-type: none"> <li>1. Biomedical sensors</li> <li>2. Smart Home sensors</li> <li>3. Roadway Safety sensors</li> <li>4. Smart Dust</li> </ol>	<ol style="list-style-type: none"> <li>1. Medical data from remote patient monitoring</li> <li>2. Home ambience settings control to save energy</li> <li>3. Road surface and Bridges monitoring for safety</li> <li>4. Motes painted on a surfaces to capture data</li> </ol>	<ol style="list-style-type: none"> <li>1. Real-time medical support from hospital staff</li> <li>2. Secured and reliable access to home owner globally</li> <li>3. Web display of potholes and ongoing constructions</li> <li>4. Repository to the collected data</li> </ol>
Low Power Personal Area Networks (6LoWPAN)	<ol style="list-style-type: none"> <li>1. Bluetooth</li> <li>2. ZigBee</li> <li>3. Body Area Network</li> <li>4. Wireless USB</li> </ol>	<ol style="list-style-type: none"> <li>1. Hands free device operations</li> <li>2. Safety and Security of Buildings</li> <li>3. Wearable technologies as smart devices</li> <li>4. Tether less connectivity to printers, scanners, etc.</li> </ol>	<ol style="list-style-type: none"> <li>1. Computational support for complex operations</li> <li>2. Data storage and communication to critical services</li> <li>3. Navigation and communications support to services</li> <li>4. Device management and security</li> </ol>
RFID Networks	<ol style="list-style-type: none"> <li>1. Merchandize tags</li> <li>2. Secured infrastructure</li> <li>3. Monitoring devices</li> <li>4. Identity devices</li> </ol>	<ol style="list-style-type: none"> <li>1. Inventory automation</li> <li>2. Access Control</li> <li>3. Wildlife monitoring</li> <li>4. Passports</li> </ol>	<ol style="list-style-type: none"> <li>1. Large data support</li> <li>2. Security policies swiftly implemented</li> <li>3. Large and possibly Global coverage</li> <li>4. Secured information access to right people</li> </ol>
Near Field Communications (NFC)	<ol style="list-style-type: none"> <li>1. Smart Transaction Devices</li> <li>2. Wireless Bootstrapping proximity devices</li> <li>3. Social Network Sharing</li> <li>4. Entertainment Devices</li> </ol>	<ol style="list-style-type: none"> <li>1. Credit Card payments</li> <li>2. Wi-Fi Direct</li> <li>3. Photos, videos, and contact sharing</li> <li>4. Mobile games</li> </ol>	<ol style="list-style-type: none"> <li>1. Security in terms of theft and fraud prevention</li> <li>2. Security of the associated devices</li> <li>3. Protecting data using secure and safe policies</li> <li>4. Provide real-time service as well as security</li> </ol>
Machine-to-Machine (M2M) and Device-to-Device (D2D) Communications	<ol style="list-style-type: none"> <li>1. Hydro Meters</li> <li>2. Factory Equipment</li> <li>3. Smart Municipalities</li> <li>4. Vehicular communication</li> </ol>	<ol style="list-style-type: none"> <li>1. Mobile users can offload tasks to cloud</li> <li>2. Automating the manufacturing process</li> <li>3. Public Utilities</li> <li>4. Safety driving, and Internet services</li> </ol>	<ol style="list-style-type: none"> <li>1. Security to prevent tampering and fraud</li> <li>2. Backend processing of large data</li> <li>3. Proper maintenance and operation of services</li> <li>4. Most economical and timely delivery of services</li> </ol>
Wireless Local Area Networks (WLAN), WiMAX, LTE/4G, and eventually 5G Networks	<ol style="list-style-type: none"> <li>1. In-home computing devices</li> <li>2. Public hotspot users</li> <li>3. Devices in airplanes and trains</li> <li>4. Devices in campuses, hotels, and businesses</li> </ol>	<ol style="list-style-type: none"> <li>1. Smartphones, Tablets, and Laptops accessing Internet</li> <li>2. Personal devices accessing data over Internet</li> <li>3. Wi-Fi coverage serving entertainment and games</li> <li>4. Internet access and data transfer between devices</li> </ol>	<ol style="list-style-type: none"> <li>1. Better access to the data from Cloud</li> <li>2. Secure delivery of data from the Cloud</li> <li>3. Global access to Cloud services</li> <li>4. Efficient and reliable access to data</li> </ol>

**Fig. 2** Network scenario for cloud based Big Data analysis for arbitrary IoT infrastructure



periodically forwards the collected data to the nearest cloud network using the front haul network. All these cloud networks are interconnected with each other by means of the fastest available communication channel links in the current technology. For example, presently single mode fiber links with petabytes/s bandwidth are available. The central cloud is a network of numerous geographically distributed cloud networks that are equipped with commercial off-the-shelf (COTS) based servers. Virtualization software based on software defined network (SDN) paradigm runs these COTS servers [30, 31]. All the network functions are realized in the software as virtual network functions (VNFs) instead of using specialized hardware to carry out switching and routing. The cloud offers these COTS servers as simple bare metal hardware services for network service provider (SP) to carry out all operations. Also, it offers VNFs as a service to the SP [32]. These SPs then offer network services to the end users, which may be a large business organization or a simple end user accessing information to acquire knowledge of a particular aspect. Such a cloud centered IoT infrastructure architecture offers solutions to many concerns raised during

handling Big Data from all these online devices [33]. The following subsections address them one by one.

### 3.1 Benefits of cloud based IoT architecture

For devices with limited energy, storage, and computation power that continuously generate data, a large repository and powerful computing power are needed to collect and process the data for daily analysis. Usually, the data must be looked from global context and sometimes correlated with other dissimilar data to observe patterns in social behavior, market performance, global warming, epidemic health crisis, and large scientific experiments. The geographically distributed IoT devices can outsource their data computation to the central cloud network for quick and efficient analysis. The output results of data analysis can be made accessible globally to the right people at the right time. The privacy of the data can be maintained using tight security measures applied to cloud networks as they are maintained by the dedicated cloud service provider (CSP) business organizations.



### 3.2 Architecture for availability, reliability, and scalability

Nowadays, high availability of online services is a main concern where cloud should offer them anytime and anywhere. IoT devices have many constraints but produce huge data collectively. For the global availability of the data, the cloud services should ensure that data is made accessible to its right users with all privileges, everywhere, and at all times. CSP business organizations determine that proper backup mechanisms can ensure that the data, analyzed results, and computation facilities (for any further analysis) are available all the time.

From a properly established and robust backup mechanism will come the next desired feature: reliability of the offered service. The services that handle and manage Big Data have to be highly reliable since many critical decisions are based on the results of Big Data analysis. New inferences and future projections will be based on the process of huge historical data. Unreliable system can mislead the study and analysis. Cloud networks offer accurate, powerful data computation, and storage infrastructure. They use software based on NFV and SDN paradigm to ensure that the network functions use the latest upgrades in the technology for reliable communication of data during distributed processing. As a result, the data analysis software can use up-to-date technologies to become more accurate and sophisticated.

Another inevitable requirement is the scalability of the cloud services. IoT infrastructure is expanding tremendously every year at an exponential rate, and the generated Big Data by this infrastructure is humongous. This problem is exacerbated with the need to store previous historical data that cannot be discarded. Therefore, a cost-effective solution that meets green communication objective is needed. Obviously, this needs constant technology upgrade with more storage capacity than the previous technology can offer. Increasing the computation power leads to power hungry technologies. Therefore, a systematic approach to the growing size of the Big Data Handling System is required. Clearly, there is no better way than moving to cloud network environment. The CSP business organizations can constantly update their technologies with greener and cost effective solutions as this is their prime business operation: that is to offer improved services to handle Big Data and its growth.

### 3.3 Green communication objective

Software domain architectures using NFV and SDN can instantiate many virtual network functions (VNF) on a single COTS server in central cloud network, where each VNF belongs to a different business end user. Traditional solutions require separate specialized hardware server

dedicated to each such business end user, which consumed more energy and space. NFV and SDN based technologies and advanced VNFs tremendously reduce communication and processor's energy consumption costs to meet green communication objectives. Moving the data processing and computation to the central cloud will substantially reduce energy bills. Further, it is easier to maintain servers in a single room or building connected to air conditioning units to reduce energy bills. IoT devices that form the end nodes consume lesser energy as they are less complex, leaving all kinds of processing and intelligence to the central cloud. The emerging cloud based radio access networks (C-RAN) are based on the same concept [34].

### 3.4 Motivation and related work

Cloud computing promises pay-as-you-go, scalable, and on-demand storage and compute services. With these promises and with fast growth of the data volume, cloud provides the best solution to store and process Big Data while minimizing or maximizing certain metrics defined in the service level agreement (SLA). Consequently, Big Data are treated as set of cloud applications with huge volume and will be scheduled accordingly on the best node in the available cloud model. In order to provide optimal deployments for data storage and processing, we propose a mixed integer linear programming (MILP) model that minimize the accompanied cost without violating the performance defined in terms of computational resources and linkage delays.

Big data is an interesting topic that has attracted many research studies. However, the way to provide performance management changes from one literature to another [35]. In [36], the authors propose a topology-aware resource allocation model. It maps the data sets, applications' VMs to servers while minimizing the execution time of the MapReduce jobs in a static cloud environment. [37] propose automated resource allocation and configuration of MapReduce environment in the cloud. Using machine-learning techniques, the model generates different Hadoop jobs' clusters and allocate them cloud resources based on the proposed optimization model.

In [38] and [39], the authors use existing Hadoop scheduling algorithms. With these algorithms, applications/data sets are scheduled in a homogeneous cloud resources environment. They do not consider any cost constraints and the impact of data analytic frequency and storage location on scheduling decision. On contrary, our work studies the impact of different factors on scheduling different Big Data sets. The proposed model minimizes the storage and processing cost while finding the optimal storage and processing location for these sets based on the analytic frequency, delay, and resources requirements.

In [40], the authors propose a model that provisions and schedules the MapReduce jobs on the available cloud nodes while minimizing the processing cost. It uses the data size and the network throughput to generate the transfer time of the data sets to the cloud and considers its impact on the cost constraints. Although the authors propose a dynamic scheduling solution while minimize the processing cost, they discard the impact of the analytic frequency on the cost calculation. Also, their SLA constraints do not consider the delay constraints between the data storage and processing location.

In [41], the authors propose a replication placement model. The latter distributes different data replicas to maximize data reliability, but it discards any cost and delay constraints. In [42], the authors minimize the communication cost of the placement of Video on Demand files while satisfying the SLA requirements based on the users' experience. [43] generates an automated data placement mechanism for cloud applications based on bandwidth cost, resource capacity constraints, and data applications interdependencies. The authors define an optimization algorithm that helps analyzing the logs submitted by the application and generating the best placement.

In [44], the authors propose a joint optimization model that minimizes the operational cost of placing different Big Data sets in multiple cloud sites. Using a two-dimensional Markov chain and non-linear optimization model, the authors distinguish multiple data processing methods and their completion time. Although their model show efficient results in terms of minimizing communication cost, the authors discard other SLA requirements impact on the placement decision such as the transmission and processing delays. Besides, the existing literature focuses only on the data processing location and the required computational resources. However, in our work, it is shown that the storage location affects the Big Data costs and processing placement. Also, the proposed solution is based on an optimization algorithm designed in form of MILP model that minimizes overall processing and storage costs while considering capacity constraints, analytic frequency effects, and network delay requirements defined in the SLA with end-users.

From the literature survey, it can be found that there are many solutions that deploy services on virtual machines in a cloud model, however, these solutions do not target Big Data and its storage and processing costs. The latter costs are evaluated based on the data size, delay, frequency of data processing, and services' renting time. In the existing solutions that discard Big Data, cost of scheduling a service is calculated in terms of delay or network traffic because the cloud provides acceptable prices when it comes to small-size data. Contrary, it is not the case when things are related to huge-size data.

This work thus focuses on providing a model for finding an optimal price and performance solutions to the Big Data processing in the cloud environment. The proposed MILP model can be used as a benchmark optimal solution based on Big Data size and location characteristics of real life IoT networks powered by huge cloud computing facilities. Since massive data storage and analysis is involved, appropriate advance reservation mechanism is assumed in the analysis. Advanced resource reservation of the resources such as link bandwidth is essential to avoid congestion in the networks while handling massive data movement during data analysis. Similar advance reservation models for cloud environment have already been proposed in literature [45–47].

#### 4 Big data management in cloud

This paper considers the central cloud network of IoT infrastructure to be a network of different cloud networks residing on different geographical locations. A typical scenario of such generalized central cloud network is shown in Fig. 3. The central cloud network consists of three different cloud networks (CN1, CN2, and CN3) connected with petabytes/s bandwidth capacity fiber optic links (L1, L2, and L3). Each cloud network is operated by different CSP; each offering different prices for storing and processing the data on their cloud servers. Three different network providers who offer different costs for transmitting data on their links operate the three fiber optic links between these cloud networks.

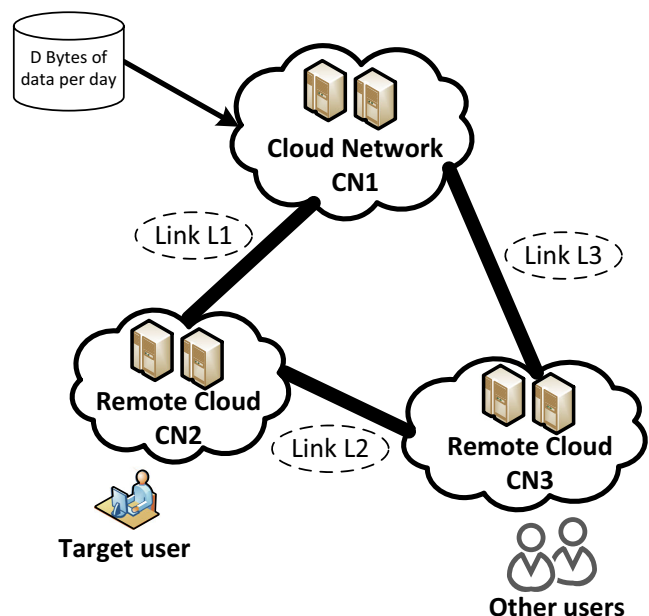


Fig. 3 Data management in cloud

For simple intuitive analysis, let us consider that the devices from IoT infrastructure directly connected to cloud network CN1 generates data of  $D$  Gigabytes every day. This data can be viewed and analyzed by the target user connected to cloud network CN2. The stored data maintains the history of 30 continuous days including the current day when data analysis is performed. Before this period, all past data is deleted. The analysis is carried only on the data collected in the past 30 days. Now, our objective is to determine where to store and process the data and how often to move it around these three cloud networks such that the overall cost of performing the data analysis is minimized. First, we calculate the cost associated with each of the different strategies or policies of storing, processing, and moving the data for analysis. Also, we show how the cost of each policy varies with the frequency of data analysis such as every day, every week, or every month (30 days).

Suppose that the costs of storing a Gigabyte of data in the server farms of cloud networks CN1, CN2, and CN3 are  $C_1^s$ ,  $C_2^s$ , and  $C_3^s$  dollars respectively. The costs of processing the same amount of data on servers of CN1, CN2, and CN3 clouds are  $C_1^p$ ,  $C_2^p$ , and  $C_3^p$  dollars respectively. Accordingly, the respective costs of transmitting a Gigabyte of data on links L1, L2, and L3 are  $C_1^l$ ,  $C_2^l$ , and  $C_3^l$  dollars. Below are the various proposed pricing policies. Although these linear models are not representative of an industry-adopted policy, but they deliberately quantify costs per gigabyte rather than time, which is subjective to the technology in place. Additionally, a similar approach has been adopted in [40] where the authors use simple and linear model to evaluate the cost of processing a data set after deploying it in a cloud environment.

### 4.1 Single-cloud, sc-policy: storing and processing on same cloud

We first consider a policy where a particular cloud from CN1, CN2, or CN3 is selected for both storing and processing the data during data analysis. Suppose that  $f$  is the frequency of data analysis. While  $f = 1$  corresponds to doing data analysis once every 30 days,  $f = 30$  corresponds to carrying out analysis every single day. In between values such as  $f = 5$  corresponds to carrying out this data analysis 5 times during a single month of 30 days.

With all of the above information in place, the overall price of performing data analysis on CN1 (*i.e.*, over a single cloud or SC-policy) is given by (SC-policy1)

$$P_1^{SC} = 30 \cdot D \cdot (C_1^s + f \cdot C_1^p) \tag{1}$$

The above equation assumes that the size of the results from data analysis is very small and insignificant. Consequently, transmission costs of the results over L1 to the target user located on CN2 is negligible.

The price of performing the same analysis through another SC-policy where CN2 is used for both storing and processing the data is given by (SC-policy2)

$$P_2^{SC} = 30 \cdot D \cdot (C_2^s + C_1^l + f \cdot C_2^p) \tag{2}$$

In the same way, if the SC-policy decides to store and process data on CN3, the overall price of performing data analysis would be (SC-policy3)

$$P_3^{SC} = 30 \cdot D \cdot (C_3^s + C_3^l + f \cdot C_3^p) \tag{3}$$

### 4.2 Multi-cloud, mc-policy: storing and processing using different clouds

It is quite possible that CSPs offer competitive prices for their services such as storage and computation to maintain their dominance in the business market and at the same time generate adequate revenue from their business to remain profitable. It is possible that the storage cost of one CSP is higher than the that of all other rival CSPs, but this particular CSP may offer the least price for its computational resources. In such a scenario, it is worth investigating to find cost saving solutions through new innovative data analysis policies that favor storing data on one cloud but use computational resources from a different cloud. This subsection investigates such multi-cloud policies (MC-policy) for performing data analysis.

Suppose that CN1 is used for data storage and CN2 is used for carrying out computational analysis (processing the data). The overall price for performing data analysis using this particular MC-policy is given by the following expression. This is labeled as MC-policy1.

$$P_{s1-p2}^{MC} = 30 \cdot D \cdot (C_1^s + f \cdot C_1^l + f \cdot C_2^p) \tag{4}$$

For the MC-policy where CN1 is used for storage and data analysis is done on CN3 (MC-policy2), the price of data analysis would be

$$P_{s1-p3}^{MC} = 30 \cdot D \cdot (C_1^s + f \cdot C_3^l + f \cdot C_3^p) \tag{5}$$

The overall price when CN3 stores data and CN1 process using MC-policy3 is

$$P_{s3-p1}^{MC} = 30 \cdot D \cdot (C_3^s + (f + 1) \cdot C_3^l + f \cdot C_1^p) \tag{6}$$

Similarly, the price when CN3 stores and CN2 processes using MC-policy4 is given by the expression as

$$P_{s3-p2}^{MC} = 30 \cdot D \cdot (C_3^s + C_3^l + f \cdot C_2^l + f \cdot C_2^p) \tag{7}$$

In the same way, the price when CN2 stores and CN1 processes using MC-policy5 is given as

$$P_{s2-p1}^{MC} = 30 \cdot D \cdot (C_2^s + (f + 1) \cdot C_1^l + f \cdot C_1^p) \tag{8}$$



Finally, the price when CN2 stores and CN3 processes using MC-policy6 is given by the following expression.

$$P_{s2-p3}^{MC} = 30 \cdot D \cdot (C_2^s + C_1^l + f \cdot C_2^l + f \cdot C_3^p) \tag{9}$$

All the above equations in this subsection provide exhaustive combinations of cloud networks using MC-policy.

### 4.3 Cost (P) versus frequency (f) of data analysis

For the sake of estimating the cost of performing data analysis each month, we have used some representative cost values for illustration as shown in Table 2. These costs are not reflective of any cost model used either in theory or practice but are useful in studying the effect of different policies on the overall cost of data analysis. We have calculated the overall cost of each data analysis policies based on the frequency of such analysis, ranging from once a month to everyday. We assume that the data generated each day, *D* is 10 Gigabytes.

In this subsection, we study the incurred cost of various SC and MC policies and see how they vary in each policy when the frequency of analysis in a month varies. These results are shown in Figs. 4 and 5 where *f* = 1 represents once a month and *f* = 30 represents carrying out data analysis everyday of the month.

From a look at Fig. 4, it can be seen that SC-policy1 that uses CN1 for storage and analysis is very expensive under lower frequency values of *f*. This is mainly because the cost of storage on CN1 is very expensive when compared with the cost of storage on CN2 and CN3. However, the CN1 offers very low computational cost. When data analysis is done more frequently, the selection of CN1 as a data analysis policy provides economical solutions. This is exactly what is seen in Fig. 4. For instance, if data analysis is done every day, the SC-policy1 offers the cheapest price. Both SC-policy2 and SC-policy3 offer similar prices that are higher compared to the price of SC-policy1. Now, it is

worth to compare the price offered by SC-policy1 with the introduced multi-cloud policies (MC-policies) to reduce the overall price further.

The price performance of different multi-cloud data analysis policies (MC-policies) with variation of the frequency of data analysis *f* is shown in Fig. 5. It can be clearly seen that MC-policy must be carefully chosen in order to reduce the price further by carefully studying the costs of resources of various service providers. If not chosen properly, the price increases instead of reducing the existing one. This can be seen clearly in Fig. 5 where not all policies offer the same price and they vary tremendously with frequency of analysis. Some are expensive while others are lower in terms price. Of all the policies, MC-policy3 offers the best price in all scenarios. This is because it combines the storage, computation, and transmissions resources in the best possible way to reduce the total cost so that the price of performing data analysis is lower all the time. This MC-policy3 is even lower compared to the best SC-policy, i.e., SC-policy1 that is identified using single-cloud based data analysis. In this section, we have considered only the cost but not the performance, and we dealt with the problem in the most simplified form. In the next section, we also consider the performance of the system while carrying out data analysis in a most cost effective way. In particular, we model the problem as an optimization model and solve it to obtain an optimal solution with regards to the selection of resources from the perspective of both price and performance.

## 5 Model for price versus performance optimization

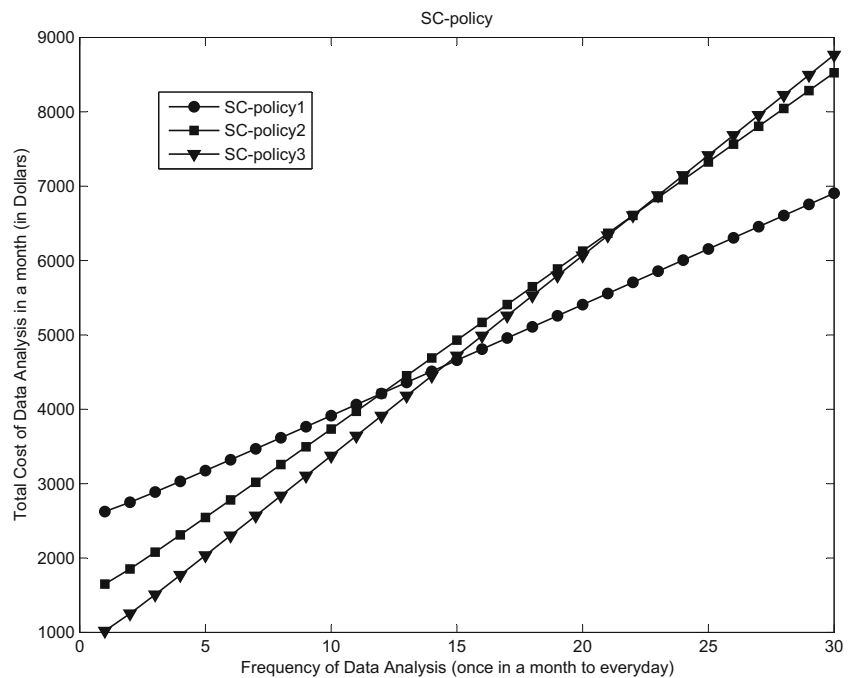
Our intuitive study reveals that appropriate handling of Big Data through an ideal policy is essential to keep the costs low while performing data analysis. An optimal way of achieving the minimal cost also needs to consider

**Table 2** Various costs involved in Data Analysis for 1 month

Costs → Cloud/Link ↓	Storage cost (dollars per Gigabyte)	Processing cost (dollars per Gigabyte)	Transmission cost (dollars per Gigabyte)
CN1	$C_1^s = 8.00$	$C_1^p = 0.50 + 0.50/(f + f^2)$	–
CN2	$C_2^s = 4.00$	$C_2^p = 0.80 + 0.50/(f + f^2)$	–
CN3	$C_3^s = 2.00$	$C_3^p = 0.90 + 0.50/(f + f^2)$	–
L1	–	–	$C_1^l = 0.40 + 0.10/(f + f^3)$
L2	–	–	$C_2^l = 0.10 + 0.10/(f + f^3)$
L3	–	–	$C_3^l = 0.20 + 0.10/(f + f^3)$

The values shown in the table are only for illustrative purposes and not representative of any charging model

**Fig. 4** The overall cost of data analysis cost as a function of frequency  $f$  using SC-policy



performance. The model proposed in this section considers all these aspects while meeting price and performance objectives. These goals are the main requirements of an end user who can be a simple service consumer or a big enterprise.

Minimizing the cost of storing and processing Big Data depends not only on the frequency of access but also on the computational and delay requirements. The latter defines the needed performance based on the SLA with end-users. Therefore, it is necessary to have an optimization model that minimizes this cost while satisfying the performance constraints. With this model, different data sets are stored and processed on same or different networks based on the following constraints:

- *Computation Resources Constraints:* These constraints ensure that the selected network should satisfy the computational requirements of a certain data set. For storing and processing purposes, the model searches for network with enough storage size in terms of terabytes, CPU cores, and power in terms of kilowatts (KW).
- *Delay Constraints:* With these constraints, the overall linkage delay  $F(L_{nn'}, L_d^u)$  between the storage and processing network should not exceed the one defined in the SLA, otherwise, user might encounter service degradation. This delay consists of both: the transmission delay that is determined using the packet length and existing transmission rate and the processing delay that depends on the distance between two cloud networks and the medium propagation speed. As for  $F_{nn'}^{av}$ , it is the existing linkage delay, which must not exceed

the required delay threshold  $F_{nn'}^{sla}$ . The latter is defined by the end-user in the SLA between the cloud provider and cloud user. Refer to Table 3 for description of all the notations used in the MILP model.

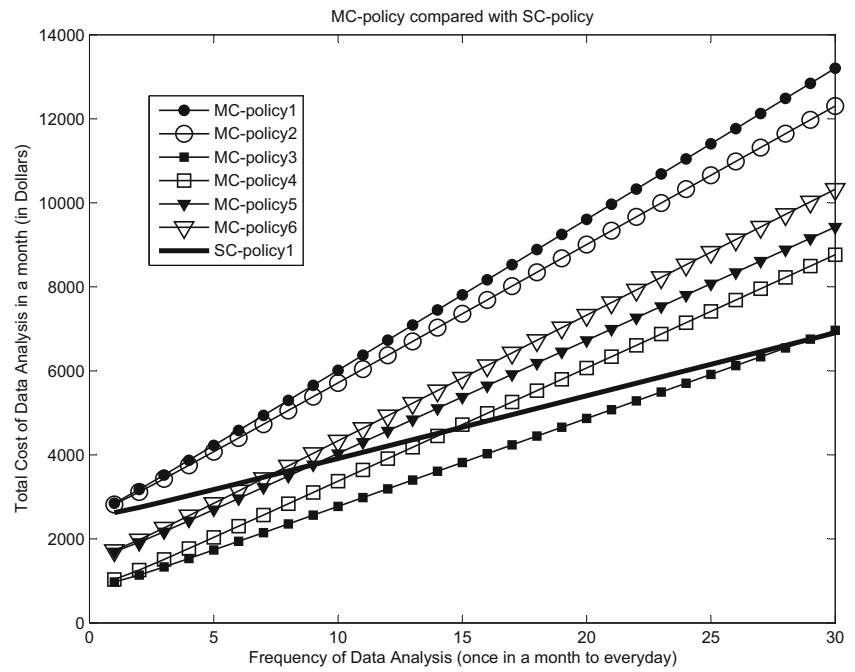
Additionally, these constraints differentiate between three different types of data: hot or frequently accessed data, warm or less-frequently accessed data, and cold or rarely accessed data. Developing such constraints necessitates indicators that refer to data type and consequently help taking best decisions regarding storing and processing that data. Therefore, we used frequency of access and propagation delay as indicators to the data type and performance. For instance, hot big data requires real-time analysis and access in order to make instant decisions when it is received. Since it is frequently accessed, the propagation delay between storage and processing networks of hot data should not exceed the threshold defined in SLA. In our case, we assume that data is managed as follows:

- Cold Data: Data accessed less than or equal to 10 days in a month.
- Warm Data: Data accessed between 11 days and 20 days in a month.
- Hot Data: Data accessed more than 21 days in a month.

### 5.1 Notations

It is assumed that the generated data is stored as one chunk on a certain network. Since we consider independent scheduled data sets, we assume that sequential processing would be best solution. This is based on the fact that

**Fig. 5** The overall cost of data analysis cost as a function of frequency  $f$  using MC-policy. The MC-policy results are compared with SC-policy to stress the importance of a proper MC-policy for data analysis



various data sources and providers indicate that sequential processing might be faster than parallel one when it comes to high data volume [48, 49]. However, parallel processing can be always adopted depending on the data type, volume, and available technologies. Moreover, this paper does not deal data analysis/processing itself. It focuses on providing

a cost effective optimal solution that is concerned with the networking aspects to support big data storage and analysis. Different parameters are used to develop the MILP model. Let  $D_i$  be the data set to be stored and processed on cloud network  $N_i$ . Table 3 shows the different notations used in the MILP model.

**Table 3** Variable Notations used in the model

Parameter	Significance
$D_i$	Data set $i$ where $i = 1, 2, 3, \dots$
$N_i$ or $CN_i$	Cloud network $i$ where $i = 1, 2, 3, \dots$
$f$	Frequency of processing certain data set in a month
$C_s$ or $C_i^s$	Storing cost of a data set on a cloud network $N_i$
$C_p$ or $C_i^p$	Processing cost of a data set on a cloud network $N_i$
$C_t$ or $C_i^t$	Transmission cost of a data set from a cloud network $N_i$ to a cloud network $N'_i$ on link $L_i$
$X_{dn}^s$	Binary decision variable that ensures if data set $d$ is stored on cloud network $n$
$Y_{dn}^p$	Binary decision variable that ensures if data set $d$ is processed on cloud network $n$
$W_{dnn'}^{sp}$	Binary decision variable that ensures if data set $d$ is stored on cloud network $n$ and processed on cloud network $n'$
$R_{dr}^s$	Required resources to store data $d$ on network $n$
$R_{dn}^s$	Available storage resources on network $n$
$R_{dr}^p$	Required resources to process data $d$ on network $n$
$R_{dn}^p$	Available processing resources on network $n$
$F_{nn'}^{av}$	Function generating the link delay between networks $n$ and $n'$
$F_{nn'}^{sla}$	Function defining the delay between networks $n$ and $n'$ to maintain certain performance baseline without violating SLA
$F(L_{nn'}, L_d^u)$	Function defining the distance between $n$ and $n'$ and the SLA distance between data storage network $n$ and data processing network $n'$

As for the decision variables, they are defined as follows:

$$X_{dn}^s = \begin{cases} 1 & \text{if } d \text{ is stored on } n \\ 0 & \text{otherwise} \end{cases}$$

$$Y_{dn}^p = \begin{cases} 1 & \text{if } d \text{ is processed on } n \\ 0 & \text{otherwise} \end{cases}$$

$$W_{dnn'}^{sp} = \begin{cases} 2 & \text{if } d \text{ is stored on } n \text{ and processed on } n' \\ 0 & \text{otherwise} \end{cases}$$

$$W_{dnn'}^{sp} = X_{dn}^s + Y_{dn'}^p \quad \forall d, n, n' \tag{17}$$

$$X_{dn}^s, Y_{dn'}^p, z_{nn'} \in \{0, 1\} \quad \forall d, n, n' \tag{18}$$

$$W_{dnn'}^{sp} \in \{0, 2\} \quad \forall d, n, n' \tag{19}$$

### 5.2 Mathematical model

The costs of storing  $C_s$ , processing  $C_p$ , and transmission  $C_t$  of certain data sets are the measures of interest of the MILP model. These three costs are related to the previously defined costs of storing  $C_i^s$  and processing  $C_i^p$  on a particular network  $N_i$  and the cost of transmitting  $C_i^l$  on link  $L_i$ . For the central cloud network scenario shown in Fig. 3, these  $C_i^s$ ,  $C_i^p$ , and  $C_i^l$  costs are already defined using Eqs. 1-9, i.e. the model considers different storing and processing policies before generating the optimal one, which might be a single cloud or multiple cloud networks. For instance,  $C_s$ ,  $C_p$ , and  $C_t$  for data set  $d_1$ , shown in Fig. 3, are written as follows:

$$C_s = 30 \cdot D \cdot (C_1^s \cdot X_{d_1n_1}^s + C_2^s \cdot X_{d_1n_2}^s + C_3^s \cdot X_{d_1n_3}^s)$$

$$C_p = 30 \cdot D \cdot f \cdot (C_1^p \cdot X_{d_1n_1}^p + C_2^p \cdot X_{d_1n_2}^p + C_3^p \cdot X_{d_1n_3}^p)$$

$$C_t = 30 \cdot D \cdot (C_3^l \cdot f \cdot W_{d_1n_1n_3}^{sp} + C_1^l \cdot f \cdot W_{d_1n_1n_2}^{sp} + \dots \dots)$$

In order to minimize  $C_s$ ,  $C_p$ , and  $C_t$ , the objective function and its constraints are formulated as follows:

$$\min \sum_d \sum_n \sum_{n'} ((C_s \times X_{dn}^s) + (C_p \times Y_{dn}^p) + (C_t \times W_{dnn'}^{sp}))$$

Subject to:

– *Computation Resources Constraints:*

$$\sum_d (X_{dn}^s \times R_{dr}^s) \leq R_{nr}^s \quad \forall n, r \tag{10}$$

$$\sum_d (Y_{dn}^p \times R_{dr}^p) \leq R_{nr}^p \quad \forall n, r \tag{11}$$

$$\sum_n X_{dn}^s = 1 \quad \forall d \tag{12}$$

$$\sum_n Y_{dn}^p = 1 \quad \forall d \tag{13}$$

$$\sum_n \sum_{n'} W_{dnn'}^{sp} = 2 \quad \forall d \tag{14}$$

– *Delay Constraints:*

$$(X_{dn}^s + Y_{dn'}^p) \leq (F_{nn'}^{sla} - F_{nn'}^{av}) \quad \forall d, n, n' \tag{15}$$

$$\begin{aligned} (X_{dn}^s \times (F(L_{nn'}, L_d^u))) &\leq h \times (1 - z_{nn'}) \\ (1 - Y_{dn'}^p) &\leq h \times z_{nn'} \quad \forall d, n, n', f \geq 20 \end{aligned} \tag{16}$$

Both transmission and propagation delays are used to calculate the overall linkage delays  $F(L_{nn'}, L_d^u)$  and  $F_{nn'}^{av}$  as follows:

$$\text{Linkage delay} = d_{trans} + d_{prop} = L/R + \mathbb{D}/s \tag{20}$$

Where  $d_{trans}$  is transmission delay,  $d_{prop}$  is propagation delay,  $L$  is the length of a packet in bits,  $R$  is the transmission rate in bits per second,  $\mathbb{D}$  is the distance between two CNs in meters, and  $s$  is the propagation speed of the media in meters per second. As for  $F_{nn'}^{av}$ , it is the required delay threshold, which must not be exceeded. It is defined by the end-user in the SLA between cloud provider and cloud user. Here any delay due to congestion is not considered as we assume appropriate reservation of transmission bandwidth resources in advance. As mentioned earlier, advance resource reservation has been proposed in the literature for cloud computing.

The proposed model minimizes the cost of storing and processing data without violating the SLA with end users. Computational resources, transmission and propagation delay, and data type constraints affect this objective. Regarding the resources constraints, Eqs. 10 and 11 ensure that the requested resources to store and process certain data set must not exceed the available resources on the selected network. Constraint (12) determines that the data can be stored on at most one network. Similarly, constraint (13) determines that the data can be processed on at most one network. As for constraint (14), it ensures that data stored on network  $n$  is transmitted to network  $n'$  and processed there. Constraint (15) ensures that the delay between the network storing Big Data and the network processing it should not violate the transmission delay requirements defined in SLA. Since hot and warm data are processed instantaneously and on demand, the propagation delay between storage and processing networks of these types should be within baseline defined in SLA. This is reflected in constraint (16). The storage, processing, and transmission costs are reflected using  $X$ ,  $Y$ , and  $W$  decision variables. When the data is stored on network  $n$  and processed on network  $n'$  then corresponding decision variables  $(X, Y) = (1, 1)$ . Consequently, the data is transmitted from the storage network to the processing one. This is shown in constraint (17). Finally, boundary constraints (18) and (19) defines binary and integer positive values for the decision variables respectively.

**Table 4** Storage and processing resources for different data sets

Required resources → Data sets ↓	Data size (Terabyte/month)	Computation resources: $R_{dr}^s$ and $R_{dr}^p$	
		CPU (vcores)	Power (kilo watts (KW))
D1	1	4	5
D2	2	6	11
D3	3	8	17

### 5.3 Performance evaluation results using MILP model

In order to study the performance results using our proposed MILP optimization model, we consider the network scenario given in Fig. 3. The network consists of two sub-models: the data set sub-model where the MILP is evaluated on 3 and 10 different data sets generated from different sources and the cloud sub-model where the data sets are distributed between three different cloud networks each having its own computational resources. With these sub-models, the optimization model minimizes the cost of storing and processing data while finding the best storage and analysis networks that satisfy the functionality constraints. However, instead of a single data set  $D$ , we consider three data sets  $D_1$ ,  $D_2$  and  $D_3$  that are injected from IoT infrastructure into cloud networks CN1, CN2, and CN3, which are represented in MILP model as  $N_1$ ,  $N_2$ , and  $N_3$  respectively. The resources required for these three data sets and the available resources on different networks are summarized in Tables 4 and 5. The computational resources are estimated based on the ones existing the market [50, 51], and [52].

All three data sets belong to different enterprise business end users and are analyzed separately. Each data set consists of different subsets coming from one data source. These subsets represent one chunk of data to be scheduled on same node. Consequently, correlated data are considered as one chunk and scheduled on same node. Ultimately, independent data sets are stored in chunks on different parts of networks because they are generated in different locations [53–55]. From networking perspective, the proposed model is conserved with how much data is moved from one location

to the other and where it is processed, all measured in the quantities of Gigabytes to avoid many technical issues and to embrace new technologies. Additionally, proper mechanisms are adhered so that all the required resources are dedicated to each data sets, and those resources are isolated from each other to avoid any security and performance compromises. We carry out data analysis on these three data set independently, from 5 times in 30 days period (a month) to every day [51, 52]. So the frequency of analysis  $f$  is varied from 5 to 30 in steps of 5.

The variation of cost in dollars for different frequency of analysis for data set  $D_1$  is shown in Fig. 6. The SLA agreement depends on the frequency of data access. Therefore, it is kept at different delay values, making it tighter as frequency of analysis is increased. It can be seen that the network offered delay is always below the corresponding SLA agreement. The result shows the minimal achievable cost in dollars while satisfying the SLA requirements, which are expressed in terms of maximum affordable delay in data analysis. The total cost represents the storage, processing, and transmission costs of data set  $d = D_1$  while the delay represents transmission and processing delays between storage network  $n$  and processing network  $n'$ .

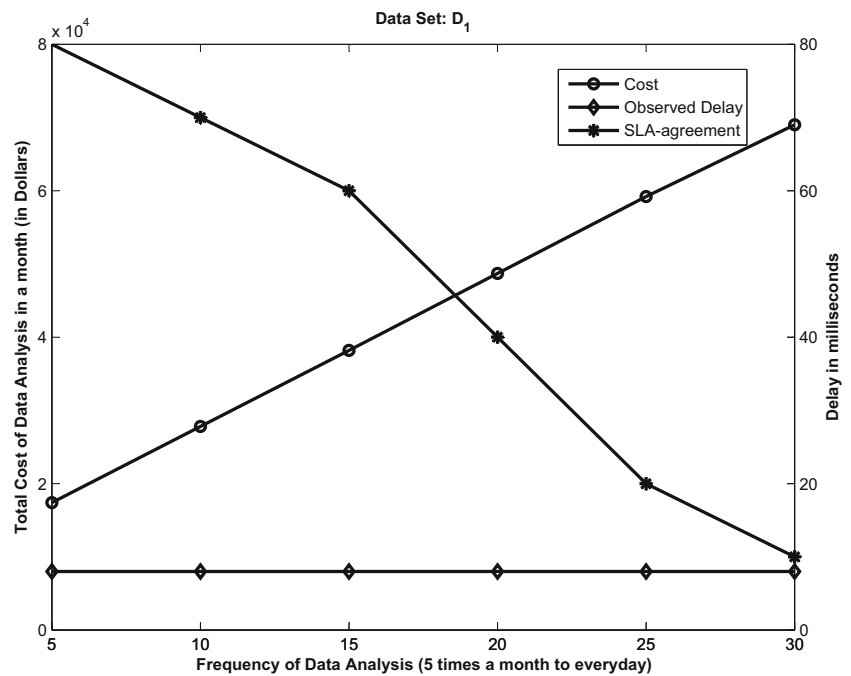
Similar performance results for data sets  $d = D_2$  and  $d = D_3$  are shown in Figs. 7 and 8 respectively. It should be noted that the amounts of resources required for these two data sets is higher than those required for the previous data set, while data set  $d = D_3$  consumes the highest resources. It can be seen from the results that the incurred costs in dollars are proportional to their consumption of resources in the network. It can be seen that for any frequency of data analysis, the SLA agreement is always satisfied. The

**Table 5** Computation resources for different networks

Available resources → Cloud networks ↓	Available storage (Exabyte)	Computation resources: $R_{dn}^s$ and $R_{dn}^p$	
		CPU (vcores)	Power (kilo watts (KW))
N1	20	16	50
N2	30	32	70
N3	25	16	60



**Fig. 6** The price versus delay performance of data set  $D_1$

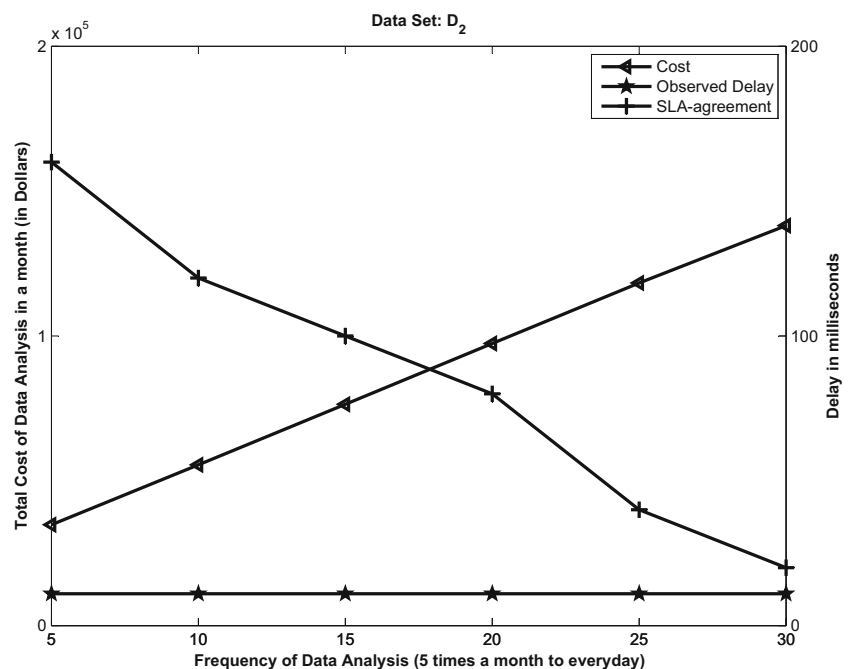


overall delay in performing data analysis is always within the affordable latency values.

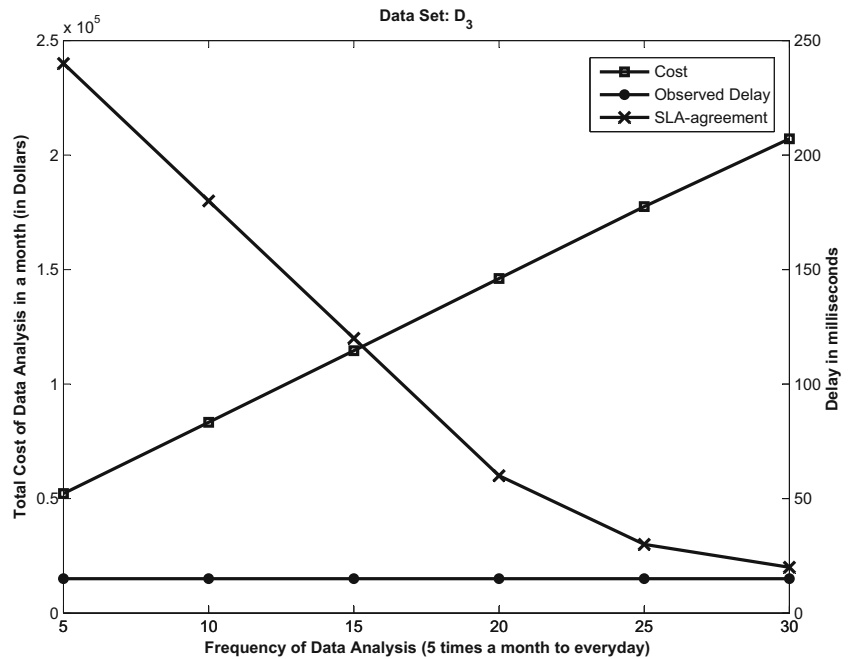
The model selects the best networks to store and process data while minimizing the associated cost. For each data set, the model is tested on different values of frequency of access. It can be concluded that the total cost increases as the data becomes frequently accessed. However, when data is analyzed and accessed frequently, propagation and transmission delays should be minimized to meet

the performance and SLA requirements. Whenever it is processed frequently, it is preferable to store and process the data on cloud networks that are close to each other. Therefore as a robust network plan, we adopt a stringent SLA policy for frequently accessed data. Due to this reason, the SLA delay becomes lower when the frequency of data access increases and consequently the SLA-agreement curve drops with increase in frequency of data access in all the presented results. It can be seen that the MILP

**Fig. 7** The price versus delay performance of data set  $D_2$



**Fig. 8** The price versus delay performance of data set  $D_3$



model always finds an optimal solution where the delay between the storage and processing networks is below the threshold delay defined in the SLA corresponding to that particular frequency of data access. This is because with the proposed MILP, the optimal placements for data processing and storing is chosen while avoiding any SLA violation. As for the observed delay generated using the MILP, it is constant and meets the performance requirements due to the static nature of the networks.

In order to extend our study to a larger network scenario, we consider 10 different data sets injected in 10 different cloud networks. All these cloud networks are connected to each other through a fully connected mesh topology of petabyte/s fiber optic links. The network capacities are different from each other but can accommodate processing of at least some of these data sets if not all of them. For instance, the storage capacity of all networks is in order of Exabyte while that of data sets requirement is only few Terabyte. Similarly the processing and power requirements of data sets are adequately met in these networks. However, all these data sets require different amounts of network resources for data analysis. In this scenario, we consider that the frequency of analysis is  $f = 15$ , which means that the analysis is done 15 times in a duration of one month (i.e., 30 days). The cost and delay performance of these 10 data sets is summarized in Table 6.

**5.4 MILP time complexity**

A scheduling problem can be defined in terms of the problem environment, problem constraints, and the

objective to be optimized. Since the proposed scheduling work has  $d$  data sets to be assigned to  $n$  cloud networks while minimizing storage, processing, and transmission cost, it can be formulated as special case of transportation problem. This case is known as the assignment problem or the bipartite matching problem. The graph has two nodes;  $n_1$  representing data sets and  $n_2$  representing cloud networks. The decision variables  $X_{dn}^s$ ,  $Y_{dn}^p$ , and  $W_{dnn'}^{sp}$  defined in Section V represents the arc that maps data sets of node  $n_1$  to networks of node  $n_2$ . This scheduling problem has NP-hard complexity hierarchy. The NP-hardness of this MILP limits its feasibility to small data set and cloud network models. In the evaluation environment, the number of variables generated in CPLEX, the optimization solver is 6916.

**Table 6** Price versus performance results of 10 data sets

Performance → Data sets ↓	Cost in dollars ( $f = 15$ )	Delay in ms
$D_1$	320000	70
$D_2$	290000	5
$D_3$	280000	30
$D_4$	190000	70
$D_5$	290000	10
$D_6$	380000	10
$D_7$	410000	60
$D_8$	280000	50
$D_9$	300000	50
$D_{10}$	300000	10

## 6 Conclusion

In this work the central network of clouds based architecture was considered for the Internet of Things (IoT) infrastructure. We have investigated that different Big Data handling policies would lead to different network costs owing to different levels of resource consumption. We have seen that cost can be minimized using multi-cloud based Big Data handling policies. This is mainly because different cloud networks provide different costs to their services offered as in many cases they are operated by different service providers under different business models. Based on these observations, we have proposed a MILP based optimization model to reduce overall costs and in particular investigate price versus performance characteristics of these networks. The MILP model minimized the storage and processing cost of a certain data set while finding the optimal location for storing and processing data and satisfying the functionality constraints defined in the SLA. These constraints included the computational resources and delay requirements. We have included up to 10 different cloud networks and 10 different data sets in terabyte sizes to carry out our study of Big Data analysis. We have seen that optimal policies in Big Data can be extended to meet green communication objectives that are vital for upcoming IoT networks including 5G. The proposed MILP optimization model can be used as a benchmark to help operators make decisions on where to store and analyze data while minimizing the accompanied cost. In future, this work will be integrated with a heuristic solution that shows its performance in more large-scale scenarios.

## References

- Mehmood Y, Ahmad F, Yaqoob I, Adnane A, Imran M, Guizani S (2017) Internet-of-things-based smart cities: recent advances and challenges. *IEEE Commun Mag* 55(9):16–24
- Xu LD, He W, Li S (2014) Internet of things in industries: a survey. *IEEE Trans Ind Inf* 10(4):2233–2243
- Cecchin C, Jimenez M, Mosser S, Riveill M (2014) An architecture to support the collection of big data in the internet of things. In: *proceedings of IEEE World Congress on Services (SERVICES)*, pp 442–449
- Lin J, Yu W, Zhang N, Yang X, Zhang H, Zhao W (2017) A survey on internet of things: architecture, enabling technologies, security and privacy, and applications. *IEEE Internet Things J* 4(5):1125–1142
- Siegel JE, Kumar S, Sarma SE (2017) The future internet of things: secure, efficient, and model-based. *IEEE Internet Things J* PP(99):1–1
- Stankovic JA (2014) Research directions for the internet of things. *IEEE Internet Things J* 1(1):3–9
- Ortiz AM, Hussein D, Park S, Han SN, Crespi N (2014) The cluster between internet of things and social networks: review and research challenges. *IEEE Internet Things J* 1(3):206–215
- Nitti M, Girau R, Atzori L (2014) Trustworthiness management in the social internet of things. *IEEE Trans Knowl Data Eng* 26(5):1253–1266
- Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M (2014) Internet of things for smart cities. *IEEE Internet Things J* 1(1):22–32
- Atzori L, Iera A, Morabito G (2011) Siot: giving a social structure to the internet of things. *IEEE Commun Lett* 15(11):1193–1195
- Liu J, Liu F, Ansari N (2014) Monitoring and analyzing big traffic data for large-scale cellular network with hadoop. *IEEE Netw* 28(4):32–39
- Hu H, Wen Y, Chua T-S, Li X (2014) Toward scalable systems for big data analytics: a technology tutorial. *IEEE Access* 2:652–687
- Ranjan R (2014) Streaming big data processing in datacenter clouds. *IEEE Cloud Comput* 1(1):78–83
- Yi X, Liu F, Liu J, Jin H (2014) Building a network highway for big data: Architecture and challenges. *IEEE Netw* 28(4):5–13
- Gurumurthi S (2009) Architecting storage for the cloud computing era. *IEEE Micro* 29(6):68–71
- Suto K, Nishiyama H, Kato N, Mizutani K, Akashi O, Takahara A (2014) An overlay-based data mining architecture tolerant to physical network disruptions. *IEEE Trans Emerg Top Comput* 2(3):292–301
- Amendola S, Lodato R, Manzari S, Occhiuzzi C, Marrocco G (2014) RFID Technology for IoT-based personal healthcare in smart spaces. *IEEE Internet Things J* 1(2):144–152
- Sheng Z, Yang S, Yu Y, Vasilakos AV, Mccann JA, Leung KK (2013) A survey on the IETF protocol suite for the internet of things: standards, challenges, and opportunities. *IEEE Wirel Commun* 20(6):91–98
- Spieß J, TJoens Y, dragnea R, Spencer P, Philippart L (2014) Using big data to improve customer experience and business performance. *Bell Labs Techn J* 18(4):3–17
- Tsai C-W, Lai C-F, Chiang M-C, Yang LT (2014) Data mining for internet of things: a survey. *IEEE Commun Surv Tutor* 16(1):77–97
- Wallace TD, Meerja KA, Shami A (2015) On-demand scheduling for concurrent multipath transfer using the stream control transmission protocol. *J Netw Comput Appl* 47:11–22
- Wallace TD, Shami A (2014) Concurrent multipath transfer using SCTP: modelling and congestion window management. *IEEE Trans Mob Comput* 13(11):2510–2523
- Dechene DJ, Shami A (2014) Energy-Aware Resource allocation strategies for LTE uplink with synchronous HARQ constraints. *IEEE Trans Mob Comput* 13(2):422–433
- Kalil M, Shami A, Al-Dweik, A (2015) QoS-Aware Power-Efficient Scheduler for LTE Uplink. *IEEE Trans Mob Comput* PP(99):1–1
- Kotval XP, Burns MJ (2013) Visualization of entities within social media: toward understanding users needs. *Bell Labs Tech J* 17(4):77–102
- Jiang L, Xu LD, Cai H, Jiang Z, Bu F, Xu B (2014) An iot-oriented data storage framework in cloud computing platform. *IEEE Trans Ind Inform* 10(2):1443–1451
- Tao F, Cheng Y, Xu LD, Zhang L, Li BH (2014) Cciot-cmfg: cloud computing and internet of things-based cloud manufacturing service system. *IEEE Trans Ind Inform* 10(2):1435–1442
- Zheng X, Martin P, Brohman K, Xu LD (2014) Cloud service negotiation in internet of things environment: a mixed approach. *IEEE Trans Ind Inform* 10(2):1506–1515
- Zheng X, Martin P, Brohman K, Xu LD (2014) Cloudqual: a quality model for cloud services. *IEEE Trans Ind Inform* 10(2):1527–1536

30. Jammal M, Singh T, Shami A, Asal R, Li Y (2014) Software-defined networking: state of the art and research challenges. *Comput Netw* 72(0):74–98
31. Hawilo H, Shami A, Mirahmadi M, Asal R (2014) Nfv: state of the art, challenges and implementation in next generation mobile networks (vepc). *IEEE Netw Mag* 28(6):18–26
32. Sharkh MA, Jammal M, Shami A, Ouda A (2013) Resource allocation in a network-based cloud computing environment: design challenges. *IEEE Commun Mag* 51(11):46–52
33. Sai V, Mickle MH (2014) Exploring energy efficient architectures in passive wireless nodes for iot applications. *IEEE Circ Syst Mag* 14(2):48–54
34. Meerja KA, Shami A, Refaey A (2015) Hailing cloud empowered radio access networks. *IEEE Wirel Commun* 22(1):122–129
35. Patikirikorala T, Colman A, Han J, Wang L (2012) A systematic survey on the design of self-adaptive software systems using control engineering approaches. In: 2012 ICSE Workshop on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), pp 33–42
36. Lee G, Tolia N, Ranganathan P, Katz RH (2011) Topology-aware resource allocation for data-intensive workloads. *SIGCOMM Comput Commun Rev* 41(1):120–124
37. Lama P, Zhou X (2012) Aroma: automated resource allocation and configuration of mapreduce environment in the cloud. In: Proceedings of the 9th International Conference on Autonomic Computing, pp 63–72
38. Verma A, Cherkasova L, Campbell RH (2011) Aria: automatic resource inference and allocation for mapreduce environments. In: Proceedings of the 8th ACM International Conference on Autonomic Computing, pp 235–244
39. Kambatla K, Pathak A, Pucha H (2009) Towards optimizing hadoop provisioning in the cloud. In: Proceedings of the 2009 Conference on Hot Topics in Cloud Computing, pp 1–5
40. Alrokayan M, Vahid Dastjerdi A, Buyya R (2014) Sla-aware provisioning and scheduling of cloud resources for big data analytics. In: 2014 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), pp 1–8
41. Cidon A, Stutsman R, Rumble S, Katti S, Ousterhout J, Rosenblum M (2013) Mincopysets: derandomizing replication in cloud storage. In: Proceedings of 10th USENIX Symposium NSDI, pp 1–5
42. Shachnai H, Tamir G, Tamir T (2012) Minimal cost reconfiguration of data placement in a storage area network. *Theor Comput Sci* 460:42–53
43. Agarwal S, Dunagan J, Jain N, Saroiu S, Wolman A, Bhogan H (2010) Volley: automated data placement for geo-distributed cloud services. In: Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation, pp 2–2
44. Gu L, Zeng D, Li P, Guo S (2014) Cost minimization for big data processing in geo-distributed data centers. *IEEE Trans Emerg Top Comput* 2(3):314–323
45. Wang W, Zhao Y, Chen H, Zhang J, Zheng H, Lin Y, Lee Y (2017) Re-provisioning of advance reservation applications in elastic optical networks. *IEEE Access* 5:10959–10967
46. Simhon E, Starobinski D (2016) A game-theoretic perspective on advance reservations. *IEEE Netw* 30(2):6–11
47. Bai H, Gu F, Shaban K, Crichigno J, Khan S, Ghani N (2015) Flexible advance reservation models for virtual network scheduling. In: 2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops), pp 651–656
48. Pfützer K Sequential processing in the age of big data, <http://www1.lehigh.edu/news/sequential-processing-age-big-data>, [Accessed: 2017]
49. TechTarget, Parallel processing in multiproviders, <http://searchsap.techtarget.com/quiz/11-Parallel-processing-in-Multiproviders>, [Accessed: 2017]
50. Feblowitz J (2012) Unleashing the power of big data in the utilities industry, Technical report. IDC Energy Insights
51. DOMO, The physical size of big data, <https://www.domo.com/learn/infographic-the-physical-size-of-big-data>, [Accessed: 2017]
52. Beal V Revolution analytics - big data analytics software, [http://www.webopedia.com/TERM/R/revolution\\_analytics\\_big\\_data\\_analytics\\_software.html](http://www.webopedia.com/TERM/R/revolution_analytics_big_data_analytics_software.html), [Accessed: 2017]
53. Wolfe J, Haghighi AD, Klein D (2008) Fully distributed em for very large datasets. In: Proceedings of the 25th International Conference on Machine Learning, pp 1–8
54. Markl V (2014) Breaking the chains: On declarative data analysis and data independence in the big data era. In: Proceedings of the VLDB Endowment, vol 7, pp 1730–1733
55. Bilmes J (2015) Summarizing large data sets, IACS Seminar Series, <http://www.seas.harvard.edu/calendar/event/81901>, [Accessed: 2015]