

# MMBD Sharing on Data Analytics Platform



Manish Devgan and Deepak Kumar Sharma

**Abstract** This chapter explores the field of Multimedia Big Data sharing on Data analytics platform. Multimedia data is a major contributor to the big data bubble. It is produced so that it can be shared among family, friends and even masses. Sharing of media data can be done in various ways and all of them have been covered in this chapter. Further, the chapter covers cloud services as a recently developed area for storage and computation. Impacts of social media giants like Facebook and Twitter along with Google Drive have been discussed. The chapter ends with a brief mention of security of online data and analysing the MMBD.

**Keywords** Big Data · Sharing · Wireless sharing · WLAN sharing · Image sharing · Cloud and benefits · Data analytics · Google cloud · Facebook · Sharing media data

## 1 Introduction

Since the mid-2000s, multimedia storing and sharing websites and services, such as Yahoo Flickr, Google photos, Dropbox, and Apple iCloud, have played a critical role in digital multimedia management. Nowadays, users have the availability of various platforms that allow them to create, manage and backup a huge amount of data. This data can be created from as simple as a smartphone to as complex as a social network such as Facebook. Facebook alone is responsible for generating 2.5 billion pieces of content and 500+ terabytes of data almost every day [1].

As multimedia data are increasing exponentially, the services that offer data sharing are providing the users with lots of options to choose from. Unlimited storage,

---

M. Devgan · D. K. Sharma (✉)  
Division of Information Technology, Netaji Subhas University of Technology, (Formerly Known as NSIT), New Delhi, India  
e-mail: [dk.sharma1982@yahoo.com](mailto:dk.sharma1982@yahoo.com)

M. Devgan  
e-mail: [manish.nsit8@gmail.com](mailto:manish.nsit8@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
S. Tanwar et al. (eds.), *Multimedia Big Data Computing for IoT Applications*, Intelligent Systems Reference Library 163,  
[https://doi.org/10.1007/978-981-13-8759-3\\_13](https://doi.org/10.1007/978-981-13-8759-3_13)

unlimited bandwidth limits, and maximum file upload sizes are just to name a few. There are services that offer a platform for sharing multimedia data to not just close-by people but also to masses. There is analytics engine that run behind these sharing services that makes ‘sense’ out of data to enhance the customer experience. Platforms could range from being a simple website or a cloud service to a more complex mobile application.

In this chapter, we will discuss on the topic of sharing multimedia data. The chapter will work on enticing the reader with the possibilities associated with the big data world. We will read about how multimedia is shared among people and will further discuss about storing this data. A brief note on security of this data will also be considered in this chapter.

We will end with discussing the Multimedia Big Data Analysis and how it helps everyone.

Finalizing the chapter will be a conclusion containing a wrap-up of the chapter.

## ***1.1 Characteristics of Big Data***

Data is, as stated above, is information that has been translated into a form that is efficient for movement or processing. Big data is characterized by the leading ‘Big’ in front of the data. This ‘big’ defines the volume of the data that has to be processed. Today, the human population generates a huge chunk of data online which is being dumped onto the servers in valleys far away from the original source of creation.

Big data doesn’t always need to be structured, i.e. be in a proper format but can also be unstructured. Here, we are only concerned with the volume of the data being produced and not the content that is being generated by a particular user or a device connected to the internet. Data, in general, saw a huge boom after the common man was introduced to the devices such as laptops and mobiles which are usually connected to the internet either wirelessly or through some wired connection. Tens of quintillion bytes of data are being generated every day and this is not just ASCII text data but also comprises of multimedia data such as audio messages, music, images, videos and even animations. This main contributor to the big data is social media platform such as Facebook (Facebook.com 2018), Snapchat (Snapchat Application, 2018) and Instagram (Instagram.com 2018).

According to recent studies and surveys around 527,760 images are shared on Snapchat every minute of the day, which is an image sharing platform service popular nowadays. More than 120 people register every minute on LinkedIn, people watch over 4 M YouTube videos every minute and post around 50 k images on Instagram, an online social media platform (Koetsier n.d.)

The above terms are enough to understand the big data term on a broad scale. But let us also see what other factors instead of volume actually define the Big Data that we are going to discuss in this chapter (Fig. 1) [2, 3].

Fig. 1 Five V's of big data



Big Data is defined using five V's [4]. They are:

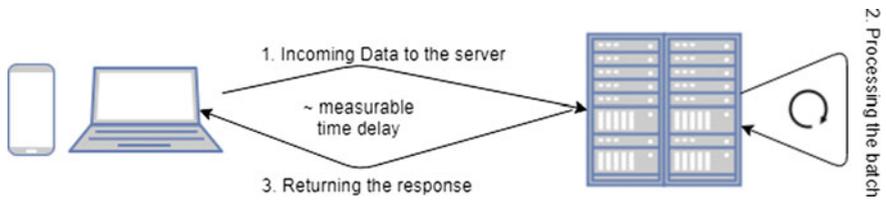
- Velocity,
- Volume,
- Value,
- Variety,
- Veracity.

Let us discuss about them in detail below [5].

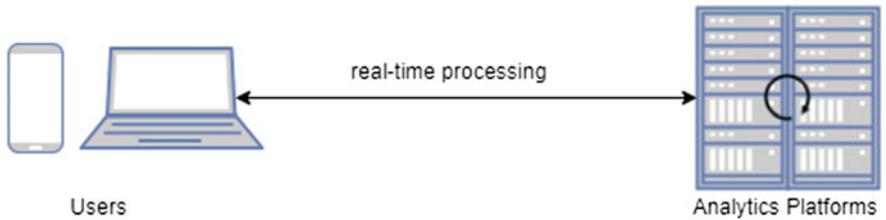
**Velocity** Velocity in big data refers to the speed at which the data is being generated. It also deals with the amount of data being processed and analysed at the same time. The emails, images, videos, animation graphics and music audio are increasing at lightning speed throughout the world, therefore, this term is everchanging as there can be no specific 'velocity' of the big data. As more and more people are being introduced to mobile and Internet technologies, the pace at which they combinedly create data over the network increases exponentially.

Initially, the services that offered analytics on big data used batch processing to handle requests. A batch would be delivered to the server to generate response and then a new batch would be formed. This was good only when the contributors to the big data bubble were not very big in number. This means when the response time was faster than the incoming data rate (Fig. 2).

Now, with increasing users the service providers have also organized themselves in a proper manner so that they can provide real-time analytics to the data being generated and stored on the platform. They use big data technologies that employ various standardized databases and API techniques that ensure that not only the



**Fig. 2** Delayed processing



**Fig. 3** Real-time processing and serving results

processing time taken for transmitting the data, from the user and back to the user is also minimum. Technologies such as Hadoop and NoSQL Databases are allowing platforms the freedom to exercise their analytics without worrying much about the time taken in storing and transmission of the same (Fig. 3).

**Volume** Volume is defined as the amount of space that a substance occupies or takes. Volume in big data is a term associated with the enormous amount of data that the machinery sensors, our mobile phones or even the simplest of reactions to any social media post contribute to this overflowing bubble of big data.

The size of consumable data is increasing day by day, for example a simple text file can range from bytes to a few kilo bytes, an image captured from a smartphone can be in a range of 2–8 MB or more, an audio file containing a song is somewhere around the same size and a movie video can be in gigabytes. There are several games that surpass the 50 GB mark as well. This shows that the current rate of data transfer and data storage on cloud is very high. This enormous data is contributing to the entropy of the big data bubble (Fig. 4).

This addition to the voluminous data is not just by active participants in the network but also by cars, credit cards, M2 M sensors, CCTV cameras and a lot more. Currently, any IOT device, i.e. network of physical objects that feature an IP address, is one of the biggest contributors to the big data. The data they produce is raw and unstructured but hold a great **value** in the data market.

Collecting and analysing this immense data is a technically difficult task, but there are certain technologies that allow us to manage them efficiently. We will discuss about them later in this chapter.

**Value** Value in terms of big data is not the worth of the raw data being stored on a platform. It is the measure of the influential data that can be generated after

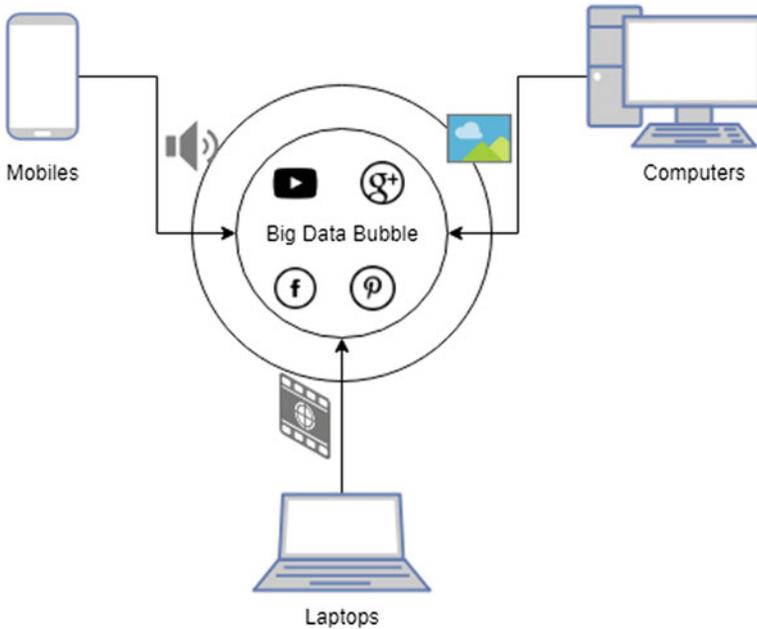


Fig. 4 Adding data to the big data bubble

processing the unstructured or raw incoming data. It is a measure of the usefulness of the processed data. While we can safely say that there is a direct link between the data and the insights of it but we cannot undermine the statement that more the data cannot always guarantee better insights, i.e. having more data does not mean having a better understanding of it. One of the most important aspects of data analysis is to understand and calculate the cost of storing, processing and analysing this data beforehand so that it does not incur any loss.

Therefore, the value associated with the data can be defined as the worth of data after *collecting, processing and analysing* and *reaping monetary benefits* out of it.

**Variety** Variety, as the name suggests, means that the incoming data being stored on the servers is not just in a single form. It need not be textual data. With the current age of modern technology and increasing numbers of media formats, we can safely assume that the big data, today, comprises mostly of multimedia data. This data can be in the form of text, audio, video, animation graphics or other formats of media storage.

Data comes in all types and formats such as structured data like numerical data in databases to unstructured data such as audio, financial market stats and more. We employ different strategies and technological hardware/software to ensure proper management of such data.

**Veracity** Veracity means ‘*conformity to the facts*’ or ‘*accuracy*’ of the facts. In big data terms, Veracity relates to biases, abnormalities and the noises in the data. It associates with the adulterated material in the big data scenario. Basically, ‘is the data being stored and analysed useful to the problem being discussed?’ is what veracity deals with. It is considered a bigger problem in data when compared to the volume or value of the stored big data.

We as analysts can try to reduce the noise as much as possible but certainly cannot reduce it to a complete zero. Data processing before analysing is a way to ensure that less amount of noise is being carried into the analysing platform and there are specific tools to identify the outliers for a specific problem so that we can have a better data to analyse.

Big data deals with issues beyond just volume and value, they are as above-mentioned veracity, variety and velocity. There is yet another V, called *volatility*, it refers to how long data is valid and must be stored before it is rendered useless for any analysing.

## 1.2 Big Data Analytics

So far, we have discussed a lot about analysing data in the above section. Here, we will discuss briefly about what data analytics is and what is its application. Analysis, in general, is a process of breaking complex things into smaller and simpler things to make a better understanding of it [6].

Use of advanced techniques such as application of complex models with elements such as predictive models, what-if analysis of data as well as applying statistical algorithms to raw data to perform analytical insights on it is what drives big data analytics. Big data analytics has given birth to various jobs such as data analysts and data scientist who analyse the ever-increasing volume of unstructured as well as structured data to ensure the growth of their respective companies or platforms.

In the next portion, we shall discuss about the applications that Big Data has in Multimedia Analysis.

### Applications of Big Data in Multimedia Analytics

Big Data techniques have been employed to store and manage the multimedia data in the industry for a good while now. This allows a sensible way to collect and store data as well as a cost-effective option to analyse and give out results. Given below are a few applications of big data in multimedia analysis.

- Social Networks

Social Networks can be considered as the paradise of Big Data. Modern day social media consists of all types of data ranging from simple text to complex animations. Social Network is a wide area of research as it houses the biggest data collection on the planet. There has been a tremendous amount of publications as well as development in the field of social media big data analysis.

Analysing human social behaviour and activities based on the Twitter Feed (Twitter.com 2018) is one of the most famous projects. The ease of availability of huge amount of hashtag and tweet data of the user allows the creation of such enormous dataset. Twitter sentiment analysis is another hot topic that data scientists are working on. Using the twitter trending to analyse the tone of a tweet. It is also known as opinion mining as it is a computational process of determining whether a particular writing is positive, negative or neutral. Another such research was conducted on Facebook's data (Facebook.com 2018) since it is considered a more valuable data source.

Another use of multimedia analytics in big data field is the emerging research field of social recommender system. Such systems can range anywhere from an online e-commerce 'What users also bought' to a video streaming platform's 'Recommended for you' section. They incorporate social data into a recommender system to get the desired outputs for a user.

- Surveillance Videos

CCTV cameras or the surveillance cameras are the biggest sources of unstructured multimedia data. They are a constant source of streaming video data that can be put to great use using the advanced analytical solutions currently available in the world. The data captured using the surveillance cameras are considered of very high value. With the upcoming big data trends and technologies, a major breakthrough can be seen in the video feed research. Object identification is one of the primary goals of video research in any surveillance camera feed. It allows us to monitor sudden and certain changes in a video by applying methodologies such as convolutional neural networks to identify objects in a live video feed.

In the US, one such project was identifying the cars number plates via the camera feeds on the traffic signals. This helped to identify and locate a lot of stolen cars and even improved the traffic laws being followed all throughout the states.

- Smart Phones

In the recent years, mobile phones particularly smart mobile phones or smartphones, in general, have taken over the electronic world. The user base of smartphone has taken over the entire userbase of laptops and PCs combined. Billions of individuals carry smartphones inside their pockets almost anywhere and everywhere. A single smartphone is equipped with technologies such as powerful CPUs, intense graphics engine, multimedia capabilities, Bluetooth, Wi-Fi and all sorts of network connection capabilities. Alongside that, the flooding of applications for smart devices has also made them a root cause of the production of vast multimedia data.

With many sorts of functionalities that a smartphone possesses there are vulnerabilities that must also be taken into account such as network security vulnerabilities that may cause the user's data into unwanted hands. We shall discuss about security later in this chapter.

- Other Areas

If we look beyond the scope of what has been discussed as the application of big data analysis in multimedia data there are several places where big data analysis is used. Multimedia summarization, Internet of Things, disaster management systems and healthcare are a few industries to name. There can be a variety of data such as records, patient's history, genomic records and a lot more that needs to be stored and analysed. Biomedicine is a growing field that has been benefited by the use of big data analytics.

## 2 Sharing Multimedia

The user's perspective of multimedia data has always been sharing it. Sharing is a primary goal of creating the multimedia data. Any sorts of multimedia are a replacement for text messages, because they can be used to express various things and not just words, they can express joy, sadness and other expressions in a single frame of a picture or multiple frames of a video.

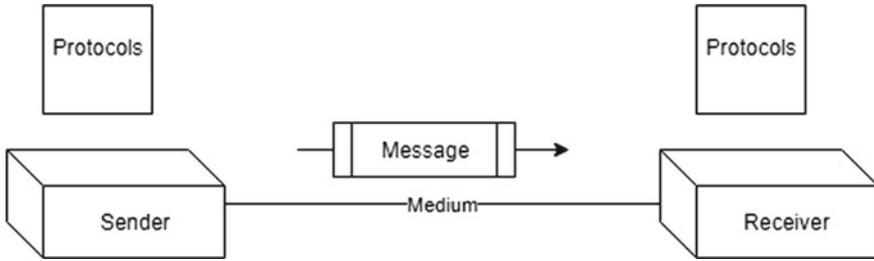
In this section, we shall study about how data is shared between devices. Though the topics are more networking based, but a brief introduction to the required subtopics has been provided.

### 2.1 *Sharing Multimedia Over a Wired Network*

This is one of the primitive ways in which data has been shared across devices. Although this is not a very old method of sharing data, there exist better methods to facilitate media sharing among devices in a network. We will discuss about the topologies that allow the formation of a network.

Wired network is an interconnection of devices that are connected to a common network through a communication medium which is wire. Two devices may be connected in some way to the same link at the same time so that communication can occur. There are two types of connections: **point-to-point (P2P)** and **multipoint**.

Figure 5 shows the involvement of a sender, a receiver, a message that contains the data that needs to be transmitted or shared, the medium is the wired connection, and some protocols governing the information exchange between the sender and receiver. The process is the same for any kind of data since the packets follow a strict protocol for being delivered.



**Fig. 5** Components of data communication

The interconnection of devices in a network is called as topology. Topology is the way in which the devices are arranged in a network connection. The way in which the network is laid out physically is defined by the topology of the network. It is a geometric representation of the links that are formed between the devices, sometimes also called as **nodes** of the network. There are four basic topologies: mesh, star, bus and ring.

There are other ways in which the nodes can be arranged in a network to create a physical layout which may be a combination of two or more of the four discussed topologies. These topologies are called **hybrid topologies**.

## 2.2 Multimedia Sharing OTA (Wirelessly)

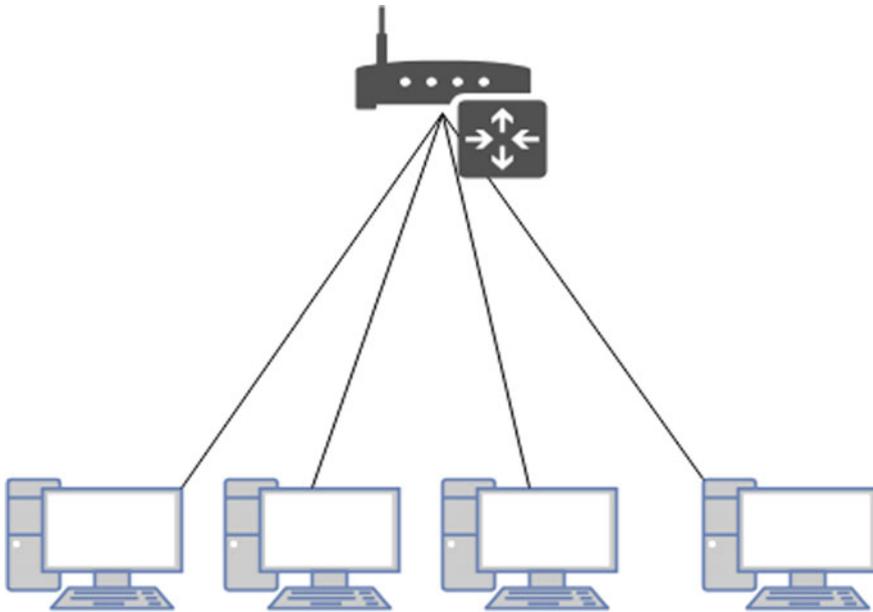
Mobile phones and laptops today come equipped with the latest technologies that allow not only sharing of data through wired medium but also sharing data via wireless mediums, for example, Bluetooth, Infrared and Wi-Fi networks [7].

In this section, we shall have a brief overview about what these technologies are and how does data transfer happen in them.

### 2.2.1 IEEE 802.11

IEEE has defined the specifications for a wireless LAN called the WLAN or the IEEE 802.11. It defines a Basic Service Set (BSS) as the building block which may or may not be controlled by a central base station also known as the access point. A BSS without an AP is an ad hoc network whereas a BSS with AP is an infrastructure network (Fig. 6).

Data transmission in a wireless network works in the same way as it works in the wired networks. The information to be transmitted is broken down into pieces called as **packets**. These packets then run through the entire network and are transmitted from the sender to the receiver. Since there are no direct physical connections between the devices and the delivery of packets must be done in order to make sure that the data



**Fig. 6** Nodes connected to a wireless router *Note: the connections are not done using wires*

**Table 1** Implementations of IEEE 802.11

FHSS	Frequency Hopping Spread Spectrum
DSSS	Direct Sequence Spread Spectrum
Infrared	Using Infrared for Transmission

is being transmitted throughout the network and any corrupt frame is retransmitted. The noisy wireless environment requires fragmentation—dividing the bigger data frames into smaller ones. It is more efficient to resend a small frame than to retransmit a large one in the entire network.

The main implementations of IEEE 802.11 WLAN are: (Table 1).

The wireless LAN technology allows the devices to share data, files, audio and a lot more on a common network without even being physically connected to the other node. The initial cost of setting up a wireless router may be high but the overall cost of maintaining the network is very less when compared to the wired network. Wireless networks also pose another advantage over the traditional wired networks that is the ease of removal and addition of a new node to the network. Previously, a new node could only be added once the central router has a free slot and a link was needed in order to connect the node to the network. This additional link added to the cost of maintaining the network. But with wireless LAN, a network is created which does not require any physical connection such as a wire. Therefore, adding a new node or removing one from the network is very easy.

### 2.2.2 Bluetooth

Bluetooth is also a wireless LAN technology designed to connect devices of different types together. The devices connected to a Bluetooth network can range from a mobile phone, headphone, notebooks, cameras and even printers. There are a lot of devices that are *Bluetooth Enabled*. A Bluetooth network is an ad hoc network which means that the network is formed spontaneously. The Bluetooth network can be connected to the internet if any one of the devices on the network has the access to the internet.

#### Applications of Bluetooth

- Connecting peripheral Bluetooth device such as wireless mouse and keyboard to communicate with the computer.
- Streaming music audio directly to the Bluetooth headset device. Bluetooth streaming of audio plays an important role in hassle-free media streaming.
- Manipulating hidden computing paradigm to make automatic synchronizations which helps in devices which carry out tasks without the user's intervention.
- Multimedia transfer is another usage of Bluetooth. Users can exchange multimedia data such as texts, images, videos and other animations using Bluetooth networks.
- Bluetooth is used by Home security devices to connect to sensors to gather data from them wirelessly.

Bluetooth, today, is an implementation of a protocol defined by the IEEE 802.15 standard. The standard defines a wireless **PAN (Personal Area Network)** which can operate in an area as small as room or maybe a hall.

Bluetooth has two network types, viz. *Piconet* and *Scatternet*.

A **piconet** is a small Bluetooth network which can have up to eight stations, one of which must be a primary station and the others will be secondary stations. There also exists a *parked state*, in which the node is in sync with the primary of the net but cannot take part in any communication or data transfer in the network.

A **scatternet** is a combination of more than one piconets in which a secondary of a piconet is the primary of other piconets (Fig. 7).

Bluetooth devices have a short-range radio transmitter. The current data rate is 1 Mbps with a 2.4 GHz bandwidth.

#### Layers of Bluetooth

Bluetooth protocol stack consists of 5-layer (Table 2).

## 2.3 Continuous Media Sharing

Continuous media sharing is a relatively new methodology that is used in streaming the content of a multimedia over the network in real time. Streaming music or video online on a service such as Spotify (Spotify.COM) or YouTube (Youtube.COM).

Media streaming solutions like these have become a solution to share data to the people in real time. One of the most popular usages of CMS is Internet Television,

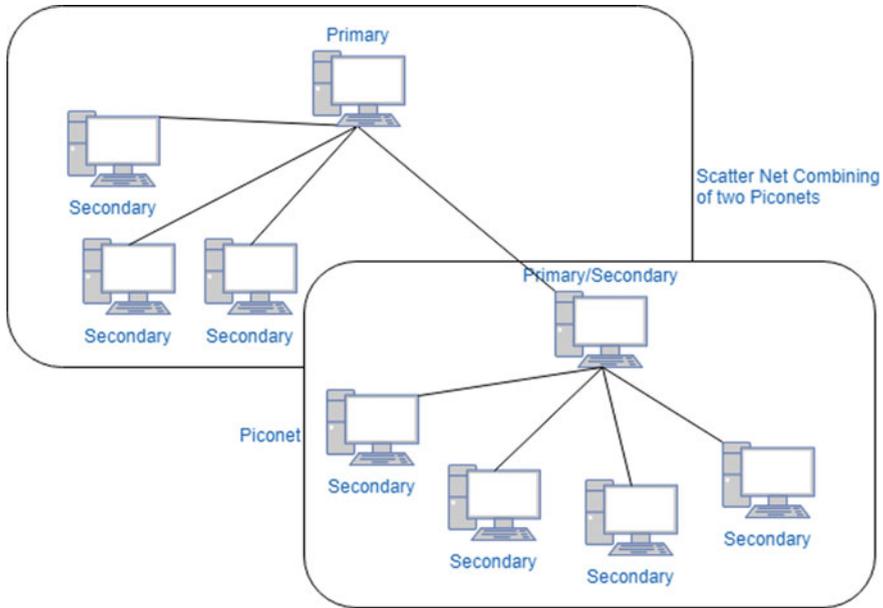


Fig. 7 Scatter-net and Piconet

Table 2 Layers of Bluetooth

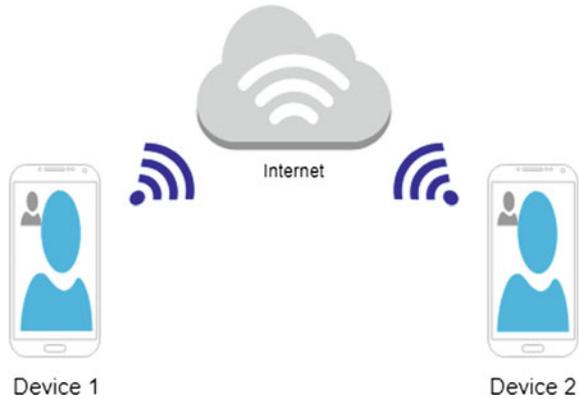
Radio	This layer specifies the requirement for radio transmission
Baseband layer	This can be considered similar to the MAC sublayer in LANs
LMP	Link Manager Protocol defines the procedure for link set up and link management
L2CAP	Logic Link Control and Adaption Protocol is responsible for adapting the upper layer protocols to the baseband layer
SDP	Service Discovery Protocol allows querying of available devices for a Bluetooth connection

providers such as Hotstar (HotStar.COM) use this service to broadcast television network over the internet to its registered users.

There are several ways in which a media can be streamed over the network.

- Serving the file over the Web.
- Using a Media Server.
- Real-Time Streaming Protocol (RTSP).

**Fig. 8** VoIP video calling



Streaming live audio and media is similar to that in case of a radio or TV, but the broadcasting is done over the internet and not the cable networks.

Real-time Transport Protocol (RTP) and RTSP are designed to handle real-time traffic on the internet and adding more functionalities to the streaming process.

**VoIP:** Voice Over IP is one such interactive audio/video application of multimedia sharing. This application allows communication over the internet to facilitate communication between two parties. Voice and Video calling support in popular chat applications such as WhatsApp and Facebook work on this principle (Fig. 8).

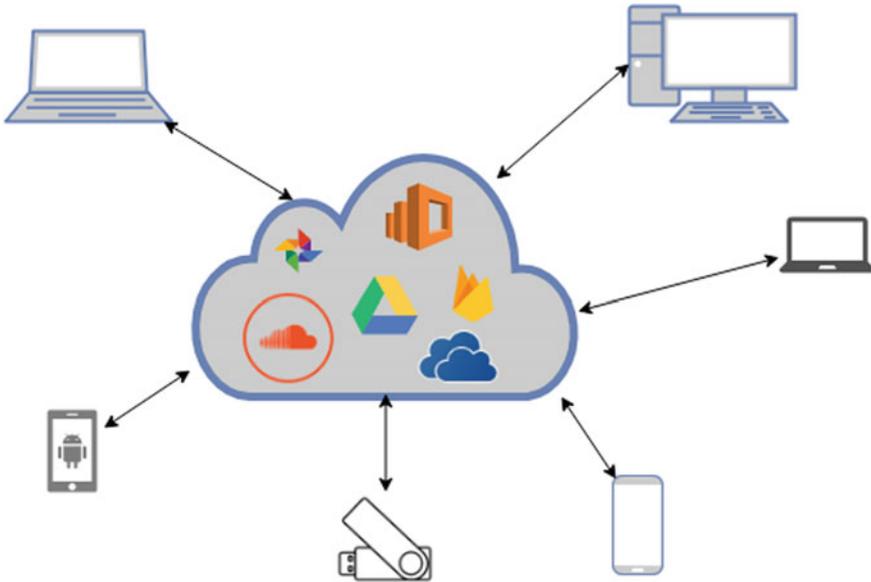
### **2.4 Mobile Networks and Cloud Sharing**

All electronic devices are capable of storing data on themselves. It can be stored in the device’s memory, often called as the internal memory or an extended service memory can be added to the device to extend the storage capacity, often a secondary memory like a MicroSD Card in a mobile phone can be used.

Now a days, people are shifting to a more cloud-based lifestyle. But What exactly is a cloud? Cloud, in simple terms is an extended memory for your device. Any device has internet connectivity can access cloud services (Fig. 9).

Data storage in cloud is done in physical or virtual servers. These servers are controlled by a Cloud Computing Provider (CCP). CCP(s) provide physical storage for use. Storage is done on physical devices in server farms. A registered user can take certain amount of storage space from the cloud for a set of pricing and use it for storing and sharing the data. Cloud services can be free up to a certain amount of storage and then charge for more or in some cases can only be used after paying a certain amount of money [8].

With internet connectivity provided by the mobile carriers, it is fairly easy to connect and use the cloud service.



**Fig. 9** Cloud services allowing seamless data transfer among all devices

### **Benefits of Cloud Services [8, 9]**

**Data storage on Cloud.** Cloud platforms provide the user with accessible storage units via the internet. These storage units service serve the purpose of storing the data online. By storing all the files and folders online, we are provided with the ease of accessibility of the file anywhere and anytime. Any electronic device can be used to access the cloud storage as long as it is connected to the internet. Using cloud solutions for managing the data enhances the productivity, operations and the efficiency of the business solution.

**Easy File Transfers.** Seamless data transfer to and from the cloud server is another intimidating feature. There are services that provide only one type of files to be stored, but mostly clouds allow heterogenous data to be stored and modified. Typically accepted files on a cloud can range from PDFs to Docs, pictures and motion videos to spreadsheets and music files. Multiple compression algorithms along with security mechanisms are used to ensure data safety as well as faster downloads and uploads.

**Backup Files on Cloud.** The most common functionality of the cloud is to backup files. Services such as Google Drive or Google Photos are constantly used to create an automatic backup of your files and media to ensure data safety in case the device is no longer available with the user. Automatic backups ensure timely updates of all files on the device.

**Sharing Media and Data.** Cloud platforms provide with different level *views*, these views allow certain restrictions on the intractability of the user. For example, uploads

can be restricted to the owner of the cloud but the files on the cloud of a user can be made available to the world for download purpose. Google Photos create a link that allows others to browse through your album without (sometimes with) the authority to add/delete media to/from it. Sharing data has been made easier with the cloud storage solutions such as OneDrive, Google Drive and Amazon Servers.

### 3 Storing and Saving Multimedia Data

Multimedia data is accompanied by a lot of **metadata**. Metadata is what defines the data. It is the data about the data generated or stored. Media data consists of a lot more than just the media, therefore, it requires more space than the media itself. Storing and serving a huge quantity of data is not easily possible and hence, provisions such as cloud storage for sharing and saving media data and other has been developed [10].

In this section, a discussion on services that offer multimedia storage has been done. Social media and its applications along with the impact that the cloud has to offer on the social media. We will also do a brief discussion on the security threats posed on the multimedia data.

#### 3.1 Services for Multimedia Storage

Multimedia storage is provided by a variety of cloud platforms. These storages support heterogenous data to be uploaded to the *drive* (*space allotted to the user is often called as a drive*). There are a lot of cloud providers, for this section we are going to compare the three biggest user-centric cloud services providers viz. Google Drive, Dropbox and OneDrive.

##### Dropbox

Dropbox is an online storage service. Dropbox is the oldest available cloud service among the three proprietary services that we are going to discuss in this chapter. It was launched to the public in 2007. It is a heterogenous cloud storage platform capable of seamlessly storing all kinds of media and data. It allows for storing of any files, sharing of the media with friends and family and an additional plus of automatically syncing local data. Although the basic services are free for Dropbox users there exists a pro version that allows for extended storage options as well. Dropbox can be availed by the regular user on all platforms like iOS, Android as well as Linux, Blackberry, etc.

##### Google Drive

In 2012, Google launched its cloud storage service called the Google Drive and hence the term *drive* was made analogous to cloud storage for general purpose users. Google

Drive made a mark after it hit the smartphones with special apps for Android, iOS Devices, etc. It operates freely with a Google account but services for more storage options can always be purchased for a varying fee. It is also a heterogenous media storage option with access to Google Drive office suite. It is available on potentially all platforms across the globe.

### **OneDrive**

OneDrive was initially launched as SkyDrive before being acquired by Microsoft and renamed to OneDrive and was opened to public in the year 2014. It is Microsoft's contender to the world of cloud storage. It is, again, a heterogenous storage service that allows all popular files to be stored on the servers. The service comes integrated with Microsoft's popular operating system Windows 8 and Windows 10. It has cross-platform applications for users using other operating systems (Table 3).

Some popular homogenous data storage services include:

**Google Photos.** Allows storing of Photos on the Cloud.

**Flickr.** Another Photo Storage Service.

**Dailymotion.** Video storage and sharing platform.

**YouTube.** Most popular public video storing and sharing platform.

## ***3.2 Social Media and Data Sharing***

Social media is a platform to connect with people. The sharing of data is not limited to text but also has images, videos and voice notes. There are various applications that are ruling the social media industry and account for the maximum media sharing across any platform.

### **3.2.1 Emergence of Facebook and Twitter**

Facebook and Twitter are the biggest names on the planet in terms of social media outage. Facebook, till date, remains the biggest social network to ever exist across the globe with a monthly usage by about 2.27 billion users. Users are allowed to share their expressions in the form of text, images and videos. Facebook and Twitter have become popular media sharing platforms used by masses to drive the world. Twitter has approximately 336 million monthly users and is more popular among movie stars and athletes as it allows a 'follow-based' system to get notified of the updates instead of the 'friend-based' system followed by Facebook.

Facebook and Twitter currently account as one of the biggest producers of multi-media data. They make sharing media easy for people. Once connected to the fellow users a user can simply 'post on the wall' or 'tweet' a sentiment, which may contain a textual data, an image or video corresponding to any topic, some graphic content and even audio files.

**Table 3** Comparison of cloud providers

Feature	Google drive	Dropbox	One drive
Ownership	Google LLC (Google.com)	Dropbox, Inc (Drobbox.com)	Microsoft (Microsoft.com)
Release	24 April 2012	June 2007	February 2014 (Following a Lawsuit)
Industry type	Cloud storage, Client software, file sharing	Online cloud, file sharing	Cloud, File share
Written in language(s)	Python, Objective-C (Mac client), wxPython (Windows)	Python, Go, Typescript	C, C++ and Python
License	Freeware	Combined with GPLv2 and proprietary software	–
Allowed free usage	15 GBs	2 GBs	5 GBs
Storage options	100 GB, 1 TB, ranging up to 30 TB	1 TB and more	50 GB, 1 TB, 5 TB
Cross-platform clients	Available on all devices like Android, Linux, Windows, iOS and more	Has applications developed for all platforms like Android, iOS, Blackberry and more along with desktop platforms	Available on all major operating systems and mobile OS(s)
Extended abilities	Yes	Yes	Yes
Abilities/Software	Google Docs, available for free to use, is an online office solution	Dropbox has Dropbox Paper, a collaborative text editor for special users	Office 365 suite
Number of users (approx.)	1 billion users	Greater than half a million	Quarter of a million users

Companies and business have been using the power of social media to improve their growth by ensuring proper content sharing across the network using dedicated Social Media teams. Twitter and Facebook have made it easier to connect to the people across the globe and hence sharing of content on a platform like this give it an edge above everything. Although both are considered equally good, but studies have shown that Facebook *posts* are 6 times more effective and have a bigger outreach than Twitter *tweet*.

### 3.2.2 Image Sharing Platforms

While social media applications have a predominant upper hand on the media sharing but after the development of dedicated media sharing applications and platforms, there has been a gradual decline in the growth of amount of data produced and shared across the two popular social network sites. Applications and platforms such as Snapchat, Instagram, Google Photos and Flickr have added to the Image sharing entropy.

**Instagram.** It is a photo and video sharing service owned by Facebook. Service allows the registered users to post images with various *captions* and *filters* to further enhance the image. It has become one of the popular choices of users to share media, specifically visible media like images and videos. Instagram has a bigger outreach to offer since most of its users are Facebook users and are quite familiar with the service. It is available for users as a cross-platform application.

**Snapchat.** Snapchat is a multimedia messaging application developed by Snap Inc. Snapchat provides the user with a completely new feature for data sharing. The pictures and messages shared across the platform were visible only for a certain amount of time and could not be retained initially and became inaccessible to the user. It has since then evolved and introduced yet another important sharing feature of *stories* of 24 h of chronological content. It is a mobile-first application and near about 200 million registered users.

**Google Photos.** While the previous two applications can be considered as a social media for image/video sharing, Google Photos is an online storage unit that allows the user to upload, download and share the clicked images and videos (only) with the people of their choice. It 'generates link' for the selected items to be shared. The visibility of the items in the album can be controlled by the original author and can also give special permission to add more content to the album etc.

**Flickr.** Flickr is an image and video hosting service acquired by Yahoo!. Developed in 2005 it has been a source of hosting images and videos for the phot researchers and bloggers across the globe.

There are other important platforms that provide media sharing capabilities like Pinterest, DeviantArt and Imgur that allow free image hosting and sharing capabilities.

## 3.3 Cloud and Its Impact

We have discussed various services that offer cloud services. Let us have an overview of the impact that the cloud has on users and businesses [11].

**Accessibility and Usability.** With the easy options of seamless transfer of files to the cloud simply by dragging and dropping files and folders, cloud provides ease of

use. Saving files on a cloud is as easy as saving them on the local machine and hence requires little to no technical knowledge.

**Disaster Recovery.** Cloud acts as a secure backup plan for business in case of an emergency. Backup is one of the most important uses that cloud storage is put to. Users can save their work, images, videos and be carefree about losing them. Snapshots and automated backups allow easy data recovery in the cloud.

**Security.** Since the data on cloud is distributed across different servers hence allows for better security of the data. Using redundant data servers along with snapshot mechanism for automating backups helps in improving security.

**Cost Saving.** Cloud is an extended storage for users as well as businesses. There is not need to be bothered about managing huge server farms for storing data performing computation on a large scale. Cloud providers provide storage units at different costs which is very less than managing the entire storage physically.

**Easy Sharing.** Cloud allows the user to share a similar environment together. The author of the original data can give permissions to the others to view, modify and delete the contents of the drive. Hence providing an easy-to-share interface.

**Automation.** Servers that store the data on cloud are configured with continuous integration or CI. CI helps in automating tasks that could be performed on the drive for different hooks. A hook is an action performed on the drive's data. In addition to it, we have periodic backups as well.

**Collaboration.** Online cloud acts as a platform for the shared learning experience. It is a collaboration platform where multiple people can access, edit, modify and collaborate on a single file. Any user with internet connectivity and access to the cloud data can collaborate with the original author.

**Scalable Service.** Service providers create plans on 'pay what you use' basis. Cloud, henceforth, is scalable in nature. The dimensionality of the environment can be changed as and when required by simply choosing a different plan or different set of configurations for the drive.

**Convenience.** Cloud provides an enriching experiencing of getting your files and folders wherever you want. The data stored on servers does not require any manual interventions as no physical device needs to be carried in order to procure the data. The seamless transfer of data from cloud to the device and vice versa makes it really convenient.

**Synchronization.** Data on the local machine is either automatically synchronized or done manually in order to update files on the cloud. It removes the stress of manually transferring the data after every update to the file or data.

Cloud provides more than just storage units, there are deployment containers that allow easy and scalable deployment environments for the builds [12]. One such free service is Heroku. There exist computationally extensive containers provided by providers such as Microsoft Azure and Amazon AWS. Google Cloud also offers computationally extensive cloud experience embedded with their own Tensorflow engines for heavy machine and deep learning.

### **3.4 *Multimedia Data and Security***

Although cloud services provide with numerous benefits, but the business enterprises still do not use the cloud for big data. It may be because of lack of visibility and not enough trust in this new infrastructure. Mobile security is another threat to multimedia big data. With the rapid escalation in the number of mobile devices, data privacy and security have been a major concern. Privacy control on different levels must be provided to ensure maximum safety of the data. Security mechanisms may include security storage and management, multi-granularity access control and privacy-aware data mining and analysis.

A multimedia platform, or a cloud platform must find the appropriate balance between access control and processing convenience. Furthermore, proper methods for encrypting multimedia data must also be employed in order to ensure data safety.

## **4 *Analysing MMBD***

### **4.1 *Data Transformations***

Data store contains the data that has been accepted for the next analysis. OLAP technology can then be used for performing this analysis. Analysis is done using data mining techniques or by the help of reporting services. This process requires a great amount of skill and expertise. It is an important step in the analysis of content and technologically heterogeneous data sources so that relevant data can be chosen. Aggregation, integration, collection and centralization of data takes place afterwards. Data pumps serve to collection and transmission of data from source systems to data stores and dumping ground. They include:

- Extraction, transmission and transformation of the data (ETL).
- Applications integration systems (EIL).

### **4.2 *Database Components***

The concept of data warehousing was coined in 1991 by Bill Inmon. Data warehouses were established as an independent information system set above business data. Data market is problem oriented, whereas, data warehouse is subject oriented. New multidimensional database models were introduced for the purpose of storing data which allowed easy and quick creation of views on data. This technology is the bases of today analytical tools of Business Intelligence (BI). Corporate Performance Management (CPM), is a new type of business planning created by the integration of BI with business planning. Data warehouses are special types of business databases

which contain consolidated data from all accessible service systems. Their optimization is done for quick administration of analytical information and not for quick data processing or quick transaction since the main aim is to mine data from the sources. They ensure processes of storing, actualization and administration of data. There exists two basic types of data stores:

#### Data Warehouse

Data warehouse is a wide centralized database for business application wherein data from all sources and external databases are saved.

#### Data Marts

Data marts are different from data warehouses as they are decentralized and thematic oriented. The analytical information that they provide is centric to one section or one specific group of people.

There exist two types of Auxiliary data stores:

Operational Data Store (ODS).

Data Staging Areas (DSA).

### **4.3 Analytical Components**

#### **OLAP [13, 14]**

Data in data warehouse is cleaned out and integrated but it is often very voluminous and hence maintaining is not easy. OLAP (Online Analytical Processing) is a special technology used for this purpose and it employs special data structures for the same. OLAP tools are simple, readily available and very popular as well as susceptible to create multidimensional analysis.

The OLAP technology is employed on multidimensional data. The data is stored in an n-dimensional cube. The database so formed is not in normalized state. A schema is formed by the tables and facts. A different visual angle is provided by every dimension on the data. Data can therefore be organized logically as well as hierarchically.

#### **Knowledge Mining from Data [15]**

##### Objective setting

It begins with a problem statement which can be related to the real world and hence begins the data mining process. The end of the process is marked by enough amount of extracted information to solve the problem statement [16]. These properties of data mining make marketing an apt area of its use.

##### Data selection

Data for data mining must be chosen carefully. In normal circumstances, data is usually extracted from source systems to a special server.

### Data preprocessing

Preparing data for analytical processing is an exacting process. It is necessary to choose corresponding information from the voluminous databases and save it to a simple relational database. Data preprocessing consist of next steps:

- Clearing Data—solving the inconsistent data or missing data problem.
- Integrating Data—multiple sources can often lead to redundant data in the server, which must be resolved.
- Data transformation—formatting of data.
- Data reduction—normalizing the data in the database as well as the formation of data models.
- Analysis and exploration of data—independent data searching without previous knowledge.
- Description—describing the complete data set.
- Prediction—Prediction phase is used to calculate the values for unknown input.

Data mining methods include

- Regression methods—linear, non-linear regressions, neural networks, etc.
- Classification of data—decision trees, SVM etc.
- Data segmentation—clustering analysis, genetic algorithms, neural clustering
- Time series prediction—Box–Jenkins method, neural networks
- Deviation detection

## 4.4 Tools for End Users [17]

### Analytical Tool—Microsoft SQL Server 2008

Just after OLAP was introduced, Microsoft began implementing a model of self-service analytical tool. In the 2005 version of Microsoft SQL Server, all analytical levels were joined into a Unified Dimension Model. MS SQL Server 2008 is the focal point in Analysis Services which are containing OLAP, Data Mining, Reporting Services and Integration Services. With the addition of SQL Server Integration Services (SSIS) it worked as an ETL data pump. It provided benefits such as allowance of creation of data administration applications. Manipulation with files in directories, data import and data export were another set of features. Reporting Services such as SQL Server Reporting Services (SSRS) provides a flexible platform for creating and distributing reports. MS SQL Server report builder is a free tool, that is usable for reports creation. SQL Server Analysis Services (SSAS) is a key analytical service for data analysis. It consists of two components:

OLAP module for multidimensional data analysis enabling loading, questioning and administration of data cubes created by Business Intelligence Development Studio (BIDS) [18].

Data Mining module which extended possibilities of business analyses.

## Data Analysis User Tools—MS Excel

Microsoft Excel is the easiest and the most obtainable of business data analytics tool. Every manager or chief executive has this software installed onto the desktop of his/her office desk. The ease and simplicity of MS Excel allow users to create reports without any external software. Graphs and reports along with other mathematical representation of data can be easily produced using MS Excel. Microsoft Excel provides a dynamic and effective range of data analysis. It offers a dynamic set of views and graphical views for the data. The most common way of inserting data to MS Excel is the manual table filling from business reports. The second way is easier and fast. It deals with importing data directly from the BIS, Business Information System. Another possible way is directly connecting the BIS to the databases. This way is most operative. MS Excel provides a set of different analysis done using pivot tables and graph pivot table. They are considered to be the most important tools of Excel. Excel enables data summarization, filtration and ordering. It is possible to create a lot of different views, reports and graphs from one data source. Creation of pivot table is easy—we can add or delete data, columns, rows or change summaries without influencing the data of other data sources. Pivot tables are very often used as a user tool for working with data cube used by SQL Server.

## 5 Conclusion

In this chapter, we discussed the field of Multimedia Big Data sharing on Data analytics platform. Multimedia data is a major contributor to the big data bubble. It is produced so that it can be shared among family, friends and even masses. Sharing of media data can be done in various ways and all of them have been covered in this chapter. Further, the chapter covered cloud services as a recently developed area for storage and computation. Impacts of social media giants like Facebook and Twitter along with Google Drive have been discussed. The chapter ends with a brief mention of security of online data and analysing the MMBD. Tools for user end analysis of data have also been given a brief mention in this chapter.

## References

1. E. Adler, Social media engagement: the surprising facts about how much time people spend on the major social networks (2016). Retrieved from <http://www.businessinsider.com/social-media-engagement-statistics-2013-12>
2. A. Labrinidis, H.V. Jagadish, Challenges and opportunities with big data. *Proc. VLDB Endow.* **5**(12), 2032–2033 (2012)
3. K. Michael, K.W. Miller, Big data: new opportunities and new challenges [guest editors' introduction]. *Computer* **46**(6), 22–24 (2013)
4. S.B. Siewert, Big data in the cloud - Data velocity, volume, variety and veracity (2013). <https://www.ibm.com/developerworks/library/bd-bigdatacloud/index.html>

5. C.L. Philip Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: a survey on big data. *Inf. Sci.* **275**(2014), 314–347 (2014)
6. P. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining* (2005). ISBN 0-321-32136-7
7. J. Bergstrom, M. Drovdaahl, S. Temple, Wireless data capture and sharing system, such as image capture and sharing of digital camera images via a wireless cellular network and related tagging of images. US Patent App. 12/182,952 (2008)
8. M.D. Assunção, R.N. Calheiros, S. Bianchi, M.A. Netto, R. Buyya, Big data computing and clouds: trends and future directions. *J. Parallel Distrib. Comput.* **79**(2015), 3–15 (2015)
9. S. Dey, A. Chakraborty, S. Naskar, P. Misra, Smart city surveillance: leveraging benefits of cloud data stores, in *Proceedings of the IEEE 37th Conference on Local Computer Networks Workshops* (IEEE, 2012), pp. 868–876
10. S. Pouyanfar, Y. Yang, S. Chen, *Multimedia Big Data Analytics: A Survey*, Florida International University (2018)
11. G. Shmueli, N. R. Patel, P. C. Bruce, *Data Mining for Business Intelligence* (2006). ISBN 0-470-08485-5
12. S. Sakr, A. Liu, D.M. Batista, M. Alomari, A survey of large scale data management approaches in cloud environments. *Commun. Surv. Tutor.* **13**(3), 311–336 (2011)
13. D. Pokorná, *Business Data Analyses Possibilities*. Diploma thesis. Faculty of Applied Informatics, Tomas Bata University in Zlín (2010)
14. M. Berthold, D. Hand, *Intelligent Data Analysis* (Springer, Berlin, 2009)
15. S.C. Hoi, J. Wang, P. Zhao, R. Jin, Online feature selection for mining big data, in *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications* (ACM, 2012), pp. 93–100
16. A. Bifet, Mining big data in real time. *Informatica (Slovenia)* **37**(1), 15–20 (2013)
17. P. Zdenka, S. Petr, S. Radek, *Data analysis: tools and methods*, Tomas Bata University in ZLín (2011)
18. H. P. Luhn, A business intelligence systems. *IBM J. Res. Dev* 314–319 (1958)