



An Introduction to Big Data and Its Possible Utility in the Urban Context

27

Sundus Samreen Wani, Salman Amin,
and Qamar Irshad

Abstract

The present-day cities are growing exponentially with urbanization and so is the population. Cities are a house to a major share of the population. With this ever-increasing population, cities come across many problems which cannot easily be anticipated. This is because, the phenomena which take place in a city are complex, interlinked, intertwined and form the city systems. In order to facilitate the solutions to the complex problems, complex data handling is required. In this digital age, big data has helped in expediting the solutions to those problems and has proven to be a boon for the complex city systems by making it easy to understand them. Big data has been put to use in various sectors in a city system; from mapping assets, solving various sectoral problems, streamlining governance and administrative issues and reducing the communication gap between stakeholders in a city. This paper aims at giving an insight about big data, and how it can be beneficial and advantageous for the city. Some successful case studies where

big data has been put to application in cities have also been incorporated into the paper.

Keywords

Big data · Digital age · City systems · Urbanization · Complex processes · Complex problems

27.1 Introduction to Big Data

ORACLE or Oak Ridge Automatic Computer and logical engine defines big data as the data that contains greater variety, arriving in increasing volumes and with more velocity. Big data are massive, complex, structured and unstructured data sets that are rapidly generated and transmitted from a wide variety of sources (Oracle 2022). It has its own characteristics, and there are parameters on the basis of which it is classified. It has its own merits and utilities. It is huge in volume, and it keeps on growing exponentially. Owing to its large, increasing size, complexity associated with it increases with time. From collection to processing to storing, the complexity remains (Taylor 2022).

One of the first examples of handling a large amount of data in modern human history could date back to the late 1800s (Kofa Study 2022; Krettek 2022; Marr 2015). This was a case of the United States census of 1890 where the estimated time to carry it out manually was estimated to be

S. S. Wani
Department of Physical Planning, School of
Planning and Architecture, New Delhi, India

S. Amin · Q. Irshad (✉)
Department of Architecture, Faculty of Architecture
& Ekistics, Jamia Millia Islamia, New Delhi, India
e-mail: qirshad@jmi.ac.in

Table 27.1 Increase in world population (in billion persons)

Year	1	1000	1500	1650	1750	1804	1850	1900	1930
Population	0.2	0.275	0.45	0.5	0.7	1	1.2	1.6	2
Year	1950	1960	1974	1980	1987	1999	2011	2020	2023
Population	2.55	3	4	4.5	5	6	7	7.8	8

1–1804 (1803 years): 0.2 to 1 billion

1804–2011 (207 years): from 1 to 7 billion

Source Worldometers.info (2022)

13 years (CrashCourse 2017). This was a predicament for the Census Bureau as the US census takes place every 10 years (United States Census Bureau 2022). The census of 1880 itself, with a resident US population of about 50 million, had taken about 7 years to compile and was published in 1888 (United States Census Bureau 2021; Anderson 2015). By the time it was completed, it had already become outdated. So, to solve this problem, the bureau had to come up with something that could handle such a large volume of data in a considerably short period of time (CrashCourse 2017).

Automation or automatic tabulation and sorting could have presented a solution. A competition was held to select a design of a machine that could best do the required job. Herman Hollerith won the competition of the US census with his tabulating machine or tabulator and sorter for its ability to count combined facts (Columbia University 2001). It used punch cards to store information of the population, and an electromechanical setup was built in the machine to read and compile the data (IBM 2022a, b). The resultant of the above was a saving of 5 million USD of 1890s taxpayers' money and compilation of the data in record time (estimates vary from 6 weeks to 3 years) (Columbia University 2001).

The US census of 1890 was a glimpse into what big data could look like, and what may be required to process it. From that time period, human civilization has only grown and grown in exponential terms. One can find examples of such growth wherever they look. The global population had taken more than 125 years to reach 2 billion in 1930 from 1 billion in 1804. The next billion was reached in 1960 after 30 years, and the next after that took only

14 years. In 1974 there were 4 billion people on the planet. So, the time period of doubling the population on the planet became 44 years from the earlier 125 years, which was a reduction, of about 80 years. In 2023, the world population is forecasted to be 8 billion as can be seen in Table 27.1. The time period of doubling may have increased to 49 years, but, where adding 1 billion people had taken 125 years, now in less than half that time, four times as many people are being added to our planet (Worldometers.info 2022). And every person comes with their own data package and information that needs to be recorded, saved and processed.

A very interesting point to note here is that the population has increased at this rate, despite the fact that the fertility rates have gone down (Roser 2014). So, the fertility rates paint a contrasting picture, and just by looking at them alone could lead one to draw a conclusion which is far from the reality. Similarly, the data increase could have pushed us to a shortage of storage, but as the amount of data increased, the capacity of our machines and computers to store it, has also gone up. On top of that, the physical sizes of our computers and the storage systems have in fact gone down. Take for example, the IBM 350 disc; it was announced in 1956, and the last model was sold till 1969. It had a storage capacity of 5 MB, while its physical size was about $1.5 \times 0.75 \times 1.75 \text{ m}^3$, which is the size of a decent-sized cloth cupboard with hangers (IBM 2022a, b; SolarWinds Worldwide 2019). Today, a data storage capacity of 128 GB and 1 TB, which are respectively 25,000 times and 200,000 times larger than that of IBM 350 disc, both take equal physical space of one fingertip each ($1.5 \times 1.1 \times 0.1 \text{ cm}^3$) (Western Digital Corporation or its affiliates 2022).

The scale of human systems continues to increase at an unprecedented rate. Through all this development of the last century, we have now reached a point where the rate of generation of information is very high, and the volume of data generated is very large. We can get an idea of how vast we have become, in terms of consuming and generating data, just by looking at the smart devices we have that continuously generate and record data—a smartphone, connected to a tablet, which is connected to a laptop; and all of these are in an ecosystem of other connected tech, such as smart TV, smart watches and smart homes. Big data is not just the information or structured or unstructured data sets, it is also the technologies, their evolution and usage that provide the required information at the correct time to the user seeking it from a sea of data that has been growing exponentially for a long time (Riahi and Riahi 2018). Big data is one the most popular phenomena of our time, that every industry wants to tap into and successfully use to fulfil their needs.

Big data classification has started with the Vs. Any data set which has these Vs are considered to be big data. ORACLE has defined three Vs of big data—volume, velocity and variety (Oracle 2022). “Volume” refers to the quantity of data collected and stored; “Velocity” refers to the rate of transfer of data; and “Variety” refers to the different types in which the data arrives, such as photographs, videos, audio and text (Arockia Panimalar et al. 2017).

Over the period of time, there was an increase in the number of Vs and they became four, then five. After further evolution, they became ten and then fourteen. The classification has now been extended to seventeen Vs and one C—volume, velocity, value, variety, veracity, validity, volatility, visualization, virality, viscosity, variability, venue, vocabulary, vagueness, verbosity, voluntariness, versatility and complexity (Arockia Panimalar et al. 2017). All of these represent a unique characteristic of big data, and further, it can be broken down, the better would its utility be.

The utility of big data presents itself in the form of big data analytics, which is the process of analysing this large volume of variable data arriving at a high velocity. Our ordinary, daily computers are

incapable of doing so, and specialized powerful machines are required for it (Riahi and Riahi 2018). Apart from being useful in general terms, big data analytics can be used specifically in an urban scenario in the following ways.

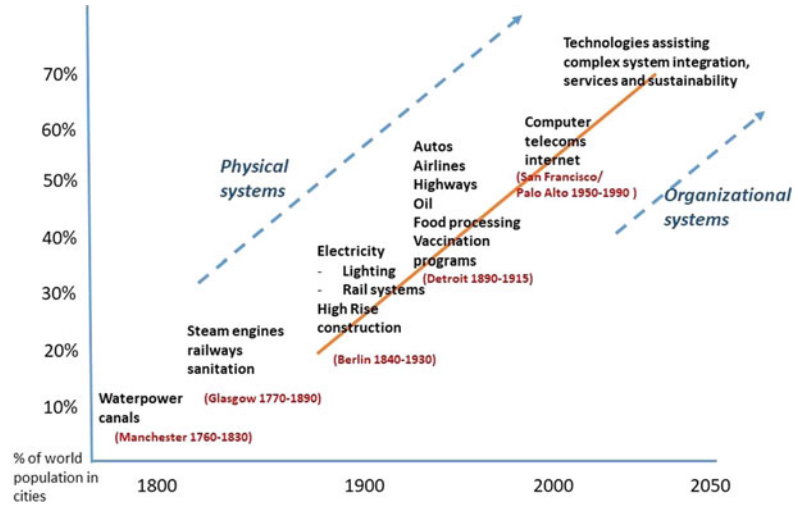
Big data analytics can be used to describe something. This is used to answer the question, “what” i.e. “what is happening” and is called descriptive analytics. Specifically in a city, it could be a phenomenon that is taking place, but is impossible to accurately describe due to the complexities in a city. Further, it can be used for diagnosis, i.e. to answer the question “why” or “why is it happening”. This is called diagnostic analytics. Similarly, there is predictive analytics, which would forecast where might we be headed or what the future of the phenomenon taking place in the city could look like. Last utility is prescriptive analytics, which recommends a path that must be taken in order to mitigate the change that the phenomenon brings with itself (Riahi and Riahi 2018).

27.2 Complexities in a City

With industrialization in the nineteenth century, urbanization became a common phenomenon. Further technological advancements, in the latter half of the twentieth century set a pace for globalization. Urbanization, technological advancement and globalization have had an impact on the social systems, built environment and the natural environment in the city, which means the systems in a city started to get altered. Technological revolution has been strongly linked to growth and development. Figure 27.1 is taken from Freeman and Perez (1988) and Hall (1999) as published in Dodgson and Gann (2011, p. 104). It shows technologies that have aided in the growth of cities along with the type of system they could be classified under or implemented with to be successful. It also includes Hall’s (1999) representative example of cities; those that have played an important role in the development of key technologies (Dodgson and Gann 2011, p. 104).

Figure 27.1 lists out technologies that developed and operated independently, over the period

Fig. 27.1 Some key technologies affecting city development and growth. Source “Freeman and Perez (1988) and Hall (1999)” in Dodgson and Gann (2011, p. 104)



of time but they are interconnected with each other. A city thus can be called as a system of systems (Dodgson and Gann 2011). For example, if we take the transportation sector into consideration, it depends upon the energy sector. Traditionally, cars were either petrol- or diesel-based, but technological advancements take place, hybrid cars that run on electricity or any other fuel got introduced. Similarly, improved telecommunication infrastructure helps in increased opportunities of working from home, which in turn has an impact on the transport sector.

Cities are complex adaptive systems which depend upon the relationships of ecosystems, organisms and the built environment. Energy and material flows through a complex network in order to get distributed. In today's world, the problems in the cities are growing and these growing problems are complex, diverse, uncertain and huge in number just like the city networks. To deal with the problems in a city, those energy and material flows need to be assessed in depth. These complex network flows are interconnected, any change in even a single component of the networks might lead to further wanted or unwanted changes in other components or networks. With the growth of cities, and more and more people settling down in the cities, metropolitan regions came up with populations in excess of 10 million people. In 1970, there were only eight megacities on the planet, by 2010, the number had grown to

27 (Kennedy et al. 2015), and a further 45 megacities likely will exist by 2030 (United Nations Department of Economic and Social Affairs Population Dynamics 2018). In 2030, 4.9 billion people will be living in the cities according to the Population Reference Bureau. Increasing size of the cities has led to increase in complexity of cities, giving rise to enormous social and environmental challenges. Cities are perceived to be areas of high global risk because of more people living in the city, with extreme levels of poverty, vulnerability and social-spatial fragmentation. To provide essential services for this increasing population, massive developments are needed as in most of the cases, these are not sustainable and do not have an equitable distribution of the resources among rich and poor. Whether cities can develop as sustainable cities depends to a large extent on how they obtain, share and manage their energy and material resources.

Problems arising due to increasing population and further developments, lead to unanticipated problems in a city. Jane Jacobs once said, “*cities happen to be problems in organized complexity*” and “People who try to predict the future by extrapolating in a line of more of what exists [today]...are always wrong”. According to him, predicting a city's future is not possible because of the complex nature of the city systems. Cities in today's world do not just face complex problems, but are stuck between wicked problems

(McFadden 2017), which are difficult or impossible to solve because of the complex interdependence (McFadden 2017). Wicked problems basically are real-world problems that acknowledge the complex interdependence of diverse factors and stakeholders, rather than simplistic, linear cause-and-effect abstractions that isolate the product of design from its context (Wahl 2017). These problems are not only limited to social systems in a city, but also to the problems associated with the reciprocal effects between social systems and natural systems that provide the basis for their existence.

If we take for instance the example of COVID-19, a very recent problem, the pandemic, one of the wicked problems that persists. It has its impact not only on the person who gets infected, but has a far-reaching multiplier impact starting from a local area level to an international level. Not only has COVID-19 impacted the health of people in various nations, but it has brought down the economy, the social status and in many cases led to pandemic-led migration. The biggest intra-country migration was seen in the Indian subcontinent since independence from the British Union in 1947 and partition of the country, when daily wage earners were forced to return to their native places from megacities due to loss of work in the sudden lockdown of 2020. An estimated 10 million workers had returned to their home state till September 2020 (Sharma 2020), i.e. within 5–6 months, which when compared to the migration at the time of Indian partition, has a higher number of migrating people per unit time. 14.5 million people had migrated over the course of 4 years at the time of independence (Bharadwaj et al. 2008). Even those who had been working formally for years in the private sector of the cities, moved back to their hometowns, in the light of lowering wages or losing jobs, the prospect of being with their family and relatives, and saving the costs pertaining to rent or higher cost of living.

The UN estimates that 55% of the global population lives in urban areas—a figure that is projected to rise to 68% by 2050. With urban sprawl, problems related to urbanization are increasing day by day. Urbanization in India has been an instrument of economic, social and

political progress but it has led to serious problems. Urban population, haphazard and unplanned growth of urban areas, and a desperate lack of infrastructure are the main problems. The rapid growth of urban population both natural and through migration has put heavy pressure on public utilities like housing, sanitation, transport, water, electricity, health, education and so on. There are visible gaps between demand and supply of facilities and services in a city, also the distribution is inequitable. A very large number of people, in cities, live in a relatively small amount of space. As the population increases in a city, areas good for development get occupied, low-income group people due to low affordability have to settle down in areas which are highly vulnerable to disasters. Intensive urban growth can lead to greater poverty, with local governments unable to provide services for all people.

Human interventions have also led to changes in the city systems, which further has altered functioning of the city. Increase in population contributes to large volumes of uncollected waste left unattended, which further creates multiple health hazards and magnifies the risk of environmental hazards. Concentrated energy use has also led to increased air pollution with significant impact on human health. By reclaiming water bodies, wetlands and other eco-sensitive areas for construction purposes, there have been changes in average temperature and rainfall. Changes in rainfall along with the changes in the land use and land cover lead to frequent floods in a city. Also, with the loss of total urban tree cover, instances of soil erosion and further turning fertile land to barren have grown across cities.

While the cities are all going through all these complex processes, the issues that we face today are difficult to anticipate and many times, cosmetic solutions that are implemented, lead to more complex and difficult situations. In order to plot the issues that we are facing in the cities today, and with the available data in a city, big data has proved to be a boon for the city planners. Ongoing digital revolution that we are going through since the last decade has been applied to the cities, and it has led to improvement in the quality of life of the people and bring

about changes in the way in which people interact and experience a city.

27.3 Application of Big Data in a City

Issues in a city are both spatial and temporal. The issues in a city are spread across either a part of a city or citywide. There is a fair enough chance that a problem that exists in a city, might not be evolving from the city itself. It might be as a result of a problem that exists in a neighbouring area, for example, the problem of air pollution in Delhi. The timing of the worst air quality of the city consistently coincides with the timing of crop burning in the neighbouring states of Haryana and Punjab. Temporal issues or problems existing in a city may vary from several minutes to years or decades. For instance, in case of a fire accident, it may last either for minutes or for hours, and this incident will be on a smaller spatial scale. Climate change on the other hand is a problem which has come into being as a result of human negligence towards the environment over a period of time. It has taken several decades to build up and reach to the point where it stands today. Generally, the more the spread of an issue is on a temporal and spatial scale, the more complex it is. Complexity in a city has been dealt with by the use of big data for efficient delivery of services and utilities. It has been used for mapping assets, solving various sectoral problems, streamlining governance and administrative issues and reducing the communication gap between stakeholders in a city. Some of the case studies where the application of big data has remained successful are as follows.

27.3.1 Santa Monica's Well-Being Project—California, USA

Well-being project is a medium to get deeper insights of community strengths and needs by collecting various data and analysing it. Data collection is done by multiple sources varying from a widespread resident survey and social

media to measure and track citizens. The data collection is dynamic so are the following processes. The findings of the data are incorporated into the formal planning process along with the budget allocation. A dynamic process makes it easy to find out measures to make the city liveable and happier continuously. This process was aimed to make a cost-effective and a streamlined process for city well-being.

Data that is collected is turned into a Well-being Index, which further is a measure of community wellbeing. This index acts as an initiation point to the understanding of community well-being and city overall. The project encourages collaboration and broad buy-in among city leaders and local organizations (e.g. needed to convince city government departments to improve their data collection) (World Economic Forum 2020).

27.3.2 Smart Dubai Happiness Meter—Dubai, UAE

This project is on similar lines as case study 1. It is developed as a part of the vision of Dubai to become the world's happiest place. As its name suggests, it measures happiness through live sentiment capture at numerous touch points across the city. City-wide technological deployment has been made for it through all possible customer interaction channels such as websites, mobile applications and physical service centres. The design used is simple in functionality to capture maximum user data. In two and a half years, more than 22 million sentiments have been captured at 4400 touchpoints via 172 customer interaction channels. Real-time data capture is helping enhance the experience of the city, along with supporting Dubai's city transformation agenda (World Economic Forum 2020).

27.3.3 Storms of the East Coast—USA

Big Data can play an important role in mitigating the harmful effects of natural disasters through disaster management. During Hurricane Sandy of

2012, the effect it was going to have on the East Coast, was studied with the help of big data analytics. The pattern of the storm was forecasted accurately, 5 days in advance, which had not been possible earlier. This enabled the authorities to take the measures necessary and plan effective responses (Simplilearn 2019).

27.4 Conclusion

Cities are a complex system of systems, and hence, the problems within the city are also complex, which are classified as wicked. To deal with the ever-challenging and exponentially rising problems, the digital age has proven to be beneficial. Rapid digitization taking place in the urban areas gives us a huge opportunity to experiment with various different methods to improve the condition of those living in the urban areas and their sprawl. Through real-time collection of data such as sentiment capturing, those who are forced to cling to the edges due to the lack of affordability, can be integrated as a stakeholder in the system, which they actually are, and be meted out equitable treatment, provided the entity in-charge has the intent in the right place. This digital age that we are going through, lends a helping hand to organize the cities by timely providing the right information to identify and mitigate the problems in a city. The problems can range from natural disasters to man-made disasters such as poor condition of living. The aforementioned right information comes from spatial and temporal checks which are necessary but dynamic, and thus need to be checked timely. In such a scenario, big data has made it possible to easily handle this data, and accurately process it, to give the results and solutions that are much required to face the ever-evolving and unpredictable, upcoming challenges.

Data storing, collection and analyses in the form of big data are definitely beneficial for any city, but the cost incurred in this whole process is an important issue. The results that we achieve are almost accurate, but at higher costs than the traditional way of data processing. Any city

administration that aims to put big data to use in the city should frame effective policies and guidelines that will help to sustain the whole process in an economical manner. If made cost-effective, big data can help to overall improve living standards and quality of life in a city overall.

References

- Anderson MJ (2015) *The American census—a social history*. Yale University Press. ISBN: 9780300195422, 0300195427
- Arockia Panimalar S., Varnekha Shree S, Veneshia Kathrine A (2017) The 17 V's of big data. *Int Res J Eng Technol (IRJET)*, 329–333
- Bharadwaj P, Khwaja AI, Mian A (2008) The big march: migratory flows after the partition of India. Faculty Research Working Papers Series. Harvard Kennedy School—John F. Kennedy School of Government
- Columbia University (2001) Herman Hollerith. Retrieved from Columbia University Computing History: <http://www.columbia.edu/cu/computinghistory/hollerith.html>
- CrashCourse (2017) Early computing: crash course computer science #1. Retrieved from YouTube: https://www.youtube.com/watch?v=O5nskjZ_GoI
- Dodgson M, Gann DM (2011) Technological innovation and complex systems in cities. *J Urban Technol* 8 (3):101–113
- Freeman C, Perez C (1988) Structural crises of adjustment, business cycles and investment behaviour. In: Dosi G, Freeman C, Nelson R, Silverberg G, Soete L (eds) *Tech Change Econ Theory* pp 38–66. London, New York: Pinter Publishers
- Hall P (1999) *Cities in Civilization: Culture, Innovation, and Urban Order*. London: Phoenix Giant
- IBM (2022a) Hollerith tabulator and sorter box. Retrieved from IBM Archives: https://www.ibm.com/ibm/history/exhibits/attic/attic_071.html
- IBM (2022b) IBM 350 disc storage unit. Retrieved from IBM Archives: https://www.ibm.com/ibm/history/exhibits/storage/storage_350.html
- Kennedy CA, Stewart I, Facchini A, Cersosimo I, Mele R, Chen B et al (2015) Energy and material flows of megacities. *PNAS* 112(19):5985–5990
- Kofa Study (2022) History of data processing (lesson 1, topic 3). Retrieved from Kofa Study—Data Processing (Week 1): <https://www.kofastudy.com/courses/jss1-computer-studies-2nd-term/lessons/data-processing-week-1/topic/history-of-data-processing/>
- Krettek A (2022) The data processing evolution: a potted history. Retrieved from ITPortal: <https://www.itportal.com/features/the-data-processing-evolution-a-potted-history/>
- Marr B (2015) A brief history of big data everyone should read. Retrieved from World Economic Forum: <https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read/>

- McFadden M (2017) Medium. Retrieved 17 Aug 2021, from <https://medium.com/@meganmcfadden/a-beginners-introduction-exploring-systems-thinking-for-21st-century-cities-faa91f4129fa>
- Oracle (2022) What is big data? Retrieved from Oracle: <https://www.oracle.com/in/big-data/what-is-big-data/>
- Riahi Y, Riahi S (2018) Big data and big data analytics: concepts, types and technologies. *Int J Res Eng*, 524–528
- Roser M (2014) Fertility rate. Retrieved from Our World in Data: <https://ourworldindata.org/fertility-rate>
- Sharma YS (2020) Labour minister Gangwar clarifies his response on migrant workers in Parliament. *The Economic Times*. Retrieved from <https://economic-times.indiatimes.com/news/economy/policy/labour-minister-gangwar-clarifies-his-response-on-migrant-labourers-in-parliament/articleshow/78142699.cms>
- Simplilearn (2019) Big data in 5 minutes | what is big data? | introduction to big data | big data explained | Simplilearn. Retrieved from YouTube: <https://www.youtube.com/watch?v=bAyrObl7TYE>
- SolarWinds Worldwide (2019) Amazing facts and figures about the evolution of hard disc drives. Retrieved from Pingdom: <https://www.pingdom.com/blog/amazing-facts-and-figures-about-the-evolution-of-hard-disc-drives/>
- Taylor D (2022) What is big data? Introduction, types, characteristics, examples. Retrieved from Guru99: <https://www.guru99.com/what-is-big-data.html>
- United Nations Department of Economic and Social Affairs Population Dynamics. (2018). World urban prospect 2018. United Nations Department of Economic and Social Affairs Population Dynamics, New York
- United States Census Bureau (2021) 1880 fast facts—history. Retrieved from U. S. Census Bureau: https://www.census.gov/history/www/through_the_decades/fast_facts/1880_fast_facts.html
- United States Census Bureau (2022) Fast facts—history—U. S. Census Bureau. Retrieved from U. S. Census Bureau: https://www.census.gov/history/www/through_the_decades/fast_facts/
- Wahl DC (2017) Facing complexity: wicked design problems. Retrieved 17 Aug 2021, from <https://medium.com/age-of-awareness/facing-complexity-wicked-design-problems-ee8c71618966>
- Western Digital Corporation or its affiliates (2022) Sandisc Extreme® microSDXC™ UHS-I CARD (Micro SD Cards). Retrieved from Western Digital: <https://www.westerndigital.com/en-in/products/memory-cards/sandisc-extreme-uhs-i-microsd#SDSQXAA-128G-GN6MA>
- World Economic Forum (2020) Smart at scale: cities to watch—25 case studies. World Economic Forum, Geneva
- Worldometers.info (2022) World population. Retrieved from Worldometer: <https://www.worldometers.info/world-population/>