Designing A Big Data Model to Improve Living Life in Smart Cities

A Dissertation

Submitted

In Partial Fulfillment of the Requirements for

The Degree of

MASTER OF TECHNOLOGY

In

Computer Science & Engineering

Submitted by:

Mohammad Sameem

Under the Supervision of:

Dr. Manish Madhava Tripathi

(Associate Professor)

Department of Computer Science & Engineering

Faculty of Engineering

INTEGRAL UNIVERSITY, LUCKNOW, INDIA

August, 2020

CERTIFICATE

This is to certify that **Mr. Mohammad Sameem** (Enroll. No. 1800103834) has carried out the research work presented in the dissertation titled **"Designing A Big Data Model to Improve Living Life in Smart Cities"** submitted for partial fulfillment for the award of the **Master of Technology in Computer Science & Engineering** from **Integral University, Lucknow** under my supervision.

It is also certified that:

(i) This dissertation embodies the original work of the candidate and has not been earlier submitted elsewhere for the award of any degree/diploma/certificate.

(ii) The candidate has worked under my supervision for the prescribed period.

(iii) The dissertation fulfills the requirements of the norms and standards prescribed by the University Grants Commission and Integral University, Lucknow, India.

(iv) No published work (figure, data, table etc.) has been reproduced in the dissertation without express permission of the copyright owner(s).

Therefore, I deem this work fit and recommend for submission for the award of the aforesaid degree.

Dr. Manish Madhava Tripathi
Dissertation Guide
(Associate Professor)
Department of CSE,
Integral University, Lucknow

Dr. Mohammadi Akheela Khanum
H.O.D.
Department of CSE ,
Integral University, Lucknow

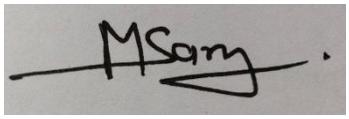
Date: 14 August, 2020
Place: Lucknow

DECLARATION

I hereby declare that the dissertation titled **"Designing A Big Data Model to Improve Living Life in Smart Cities"** is an authentic record of the research work carried out by me under the supervision of Dr. Manish Madhava Tripathi, Department of Computer Science & Engineering , for the period from August ,2019 to August , 2020 at Integral University, Lucknow. No part of this dissertation has been presented elsewhere for any other degree or diploma earlier.

I declare that I have faithfully acknowledged and referred to the works of other researchers wherever their published works have been cited in the dissertation. I further certify that I have not willfully taken other's work, para, text, data, results, tables, figures etc. reported in the journals, books, magazines, reports, dissertations, theses, etc., or available at web-sites without their permission, and have not included those in this M.Tech dissertation citing as my own work.

Date: 14 August, 2020



Signature

Mohammad Sameem
Enroll. No.1800103834

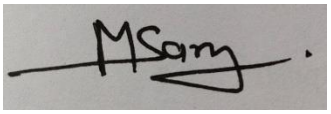
COPYRIGHT TRANSFER CERTIFICATE

Title of the Dissertation: **Designing A Big Data Model to Improve Living Life in Smart Cities.**

Candidate Name: **Mohammad Sameem**

The undersigned hereby assigns to Integral University all rights under copyright that may exist in and for the above dissertation, authored by the undersigned and submitted to the University for the Award of the M.Tech degree.

The Candidate may reproduce or authorize others to reproduce material extracted verbatim from the dissertation or derivative of the dissertation for personal and/or publication purpose(s) provided that the source and the University's copyright notices are indicated.



Mohammad Sameem

ACKNOWLEDGEMENT

I am highly grateful to the Head of Department of Computer Science and Engineering for giving me proper guidance and advice and facility for the successful completion of my dissertation.

It gives me a great pleasure to express my deep sense of gratitude and indebtedness to my guide **Dr. Manish Madhava Tripathi, Associate Professor, Department of Computer Science and Engineering,** for his valuable support and encouraging mentality throughout the project. I am highly obliged to him for providing me this opportunity to carry out the ideas and work during my project period and helping me to gain the successful completion of my Project.

I am also highly obliged to **Dr. Mohammadi Akheela Khanum (Associate Professor, Department Of Computer Science and Engineering)** and PG Program Coordinator **Dr. Faiyaz Ahamad, Assistant Professor, Department of Computer Science and Engineering,** for providing me all the facilities in all activities and for his support and valuable encouragement throughout my project.

My special thanks are going to all of the faculties for encouraging me constantly to work hard in this project. I pay my respect and love to my parents and all other family members and friends for their help and encouragement throughout this course of project work.

Date: 14 August, 2020
Place: Lucknow

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

SC	Smart City
IoT	Internet of Things
ETL	Extract, Transform & Load
AI	Artificial Intelligence
HDFS	Hadoop Distributed System
SQL	Structured Query Language
HSQL	Hive Structured Query Language
PDW	Parallel Data Warehouse
GDPR	General Data Protection Regulation
ITS	Intelligent Transportation System
GPS	Global Positioning System
V2V	Vehicle-to-vehicle
V2I	Vehicle-to-Infrastructure

ABSTRACT

Welcome to the era of either big data, as we all know this is the era of big data and every model in this era is directly or indirectly by utilizing the concepts of big data. Most important aspect of big data is to handle volume, velocity and Variety. While implementing the concept of smart city these features of big data will play an important role to efficiently manage various services. Transportation can be handily overseen with the aid of Big Data. A jam or a congestion on road is a very weary thing for commuters. The analysis of vehicular movement related data work can be done to address this issue effectively. Big Data will also serve this purpose. In our report we are going to prepare, a data analytics-based model to tackle all these problems. For this we will utilize the techniques of big data because traffic related data segment in any urban city has voluminous in nature and size of data will decrease as traffic density will increase. The aim of this report is to estimate the conduct of the urban traffic of any urban city. The strategy of the report comprises in the catch of important occasions that influence the progression of traffic of the city and the utilization of information investigation to prepare with these prominent events to foresee the conduct of traffic. This report presents the outcomes in anticipating the conduct of the traffic of the city for seven days.

CHAPTER: 1

INTRODUCTION

1.1 INTRODUCTION OF BIG DATA

Big Data is additionally data but with an enormous size. Big Data may be a term won't to describe a set of knowledge that's huge in volume and yet growing exponentially with time. In short, such data is so large and sophisticated that none of the normal data management tools are ready to store it or process it efficiently.

Simply, big data is very big, more complex data, especially from new data sources. These data sets are so voluminous that traditional processing software just can't manage them. But these massive volumes of knowledge are often wont to address business problems you wouldn't are ready to tackle before.

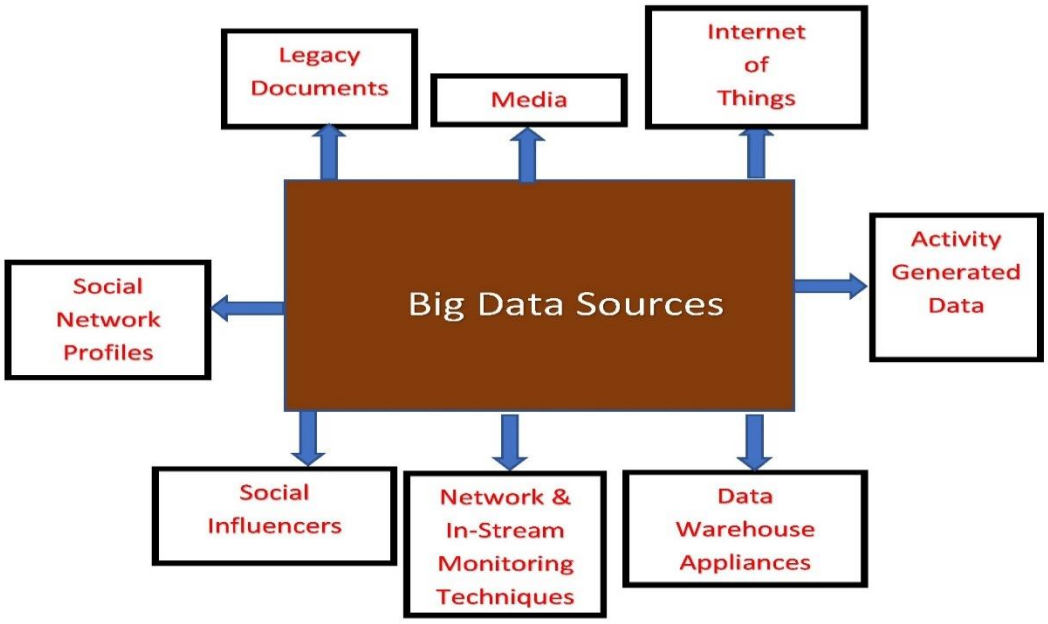


Figure 1.1: Big Data Sources

In big data, data contains greater variety outcomes in rapidly increasing volumes and with higher velocity. This is known as the three Vs.

1.1.1 THE THREE Vs OF BIG DATA

Volume – The name Big Data itself is said to a size which is gigantic. Size of data plays a very crucial role in determining value out of data. Also, whether a specific data can actually be considered as an enormous Data or not, depends upon the quantity of knowledge. Hence, 'Volume' is one characteristic which must be considered while handling Big Data.

Velocity – The term 'velocity' refers to the speed of generation of knowledge. How fast the info is generated and processed to satisfy the stress, determines real potential within the data. Velocity of data in big data means the speed at which data flows in from of sources like businesses, applications, network areas, sensors, mobile devices, etc. The flow of data is massive and continuous.

Variability – It refers to the inconsistency which can be shown by the data at times, so hampering the process of being able to handle and manage the data effectively and easily.

There are two more Vs have emerged over the past few years: Value and Veracity.

Data has intrinsic value. But it's of no use until that value is known and knowledgeable. Equally important: How truthful is your data—and how much can you rely on it?

Today, big data has become capital. Think of some of the world's biggest tech companies. A large and important part of the value they offer comes from their business data, which they're constantly analyzing to produce more efficiency and develop new products.

With an increased volume of massive data now cheaper and more accessible, you'll make more accurate and precise business decisions.

Finding value in big data isn’t only about analyzing it (which is a whole other benefit). It’s an entire discovery process that requires insightful analysts, business users, and executives who ask the right questions, recognize patterns, make informed assumptions, and predict behavior.

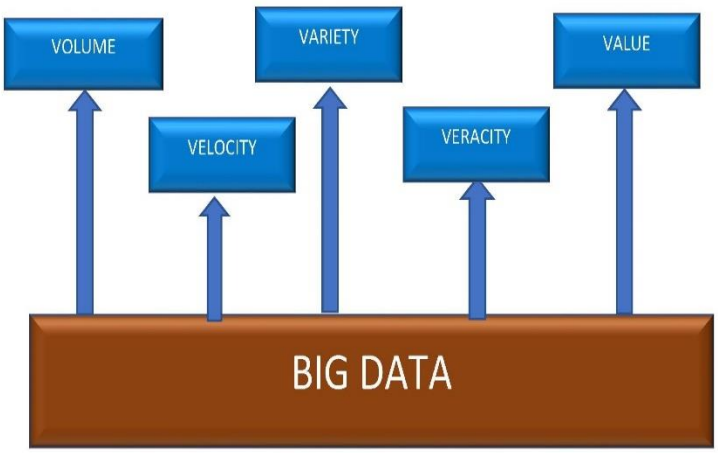


Figure 1.2: The Three Vs of Big Data

1.1.2 THE HISTORY OF BIG DATA

Although the concept of massive data itself is comparatively new, the origins of huge data sets return to the 1960s and '70s when the planet of knowledge was just getting started with the first data centers and the development of the relational database. Around 2005, people began to understand just what proportion data users generated through Facebook, YouTube, and other online services. Hadoop (an open-source framework created specifically to store and analyses big data sets) was developed that same year. NoSQL also began to realize popularity during this point.

The development of open-source frameworks, like Hadoop (and more recently, Spark) was essential for the expansion of massive data because they create big data easier to figure with and cheaper to store. In the years since then, the quantity of massive data has skyrocketed. Users are still generating huge amounts of data—but it's not just humans who do it. With the advent of the Internet of Things (IoT), more objects and devices are connected to the internet, gathering data on customer usage patterns and product performance. The emergence of machine learning has produced still more data.

While big data has come far, its usefulness is only just beginning. Even further with cloud computing expanded possibilities of big data. The cloud offers truly elastic scalability, where developers can simply spin up unplanned clusters to check a subset of knowledge.

1.1.3 PARTS OF BIG DATA

Big Data' might be found in three forms:

Structured

Any data which will be stored, accessed and processed within the sort of fixed format is termed as a 'structured' data. Over the amount of your time, talent in computing has achieved greater success in developing techniques for working with such quite data (where the format is documented in advance) and also deriving value out of it. However, nowadays, we are foreseeing issues when a size of such data grows to an enormous extent, typical sizes are being within the rage of multiple zettabytes.

Unstructured

Any data with unknown form or the structure is assessed as unstructured data. additionally, to the dimensions being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it. A typical example of unstructured data may be a heterogeneous data source containing a mixture of straightforward text files, images, videos etc. Now day organizations have wealth of knowledge available with them but unfortunately, they do not skills to derive value out of it since this data is in its raw form or unstructured format.

Semi-Structured

Semi-Structured data can contain both the sorts of data. we will see semi-structured data as a structured in form but it's actually not defined with e.g. a table definition in relational DBMS. Example of semi-structured data may be a data represented in an XML file.

1.1.4 BENEFITS OF MASSIVE DATA AND DATA ANALYTICS

• Big data makes it possible for you to realize more complete answers because you've got more information.

• More complete answers mean more confidence within the data—which means a totally different approach to tackling problems.

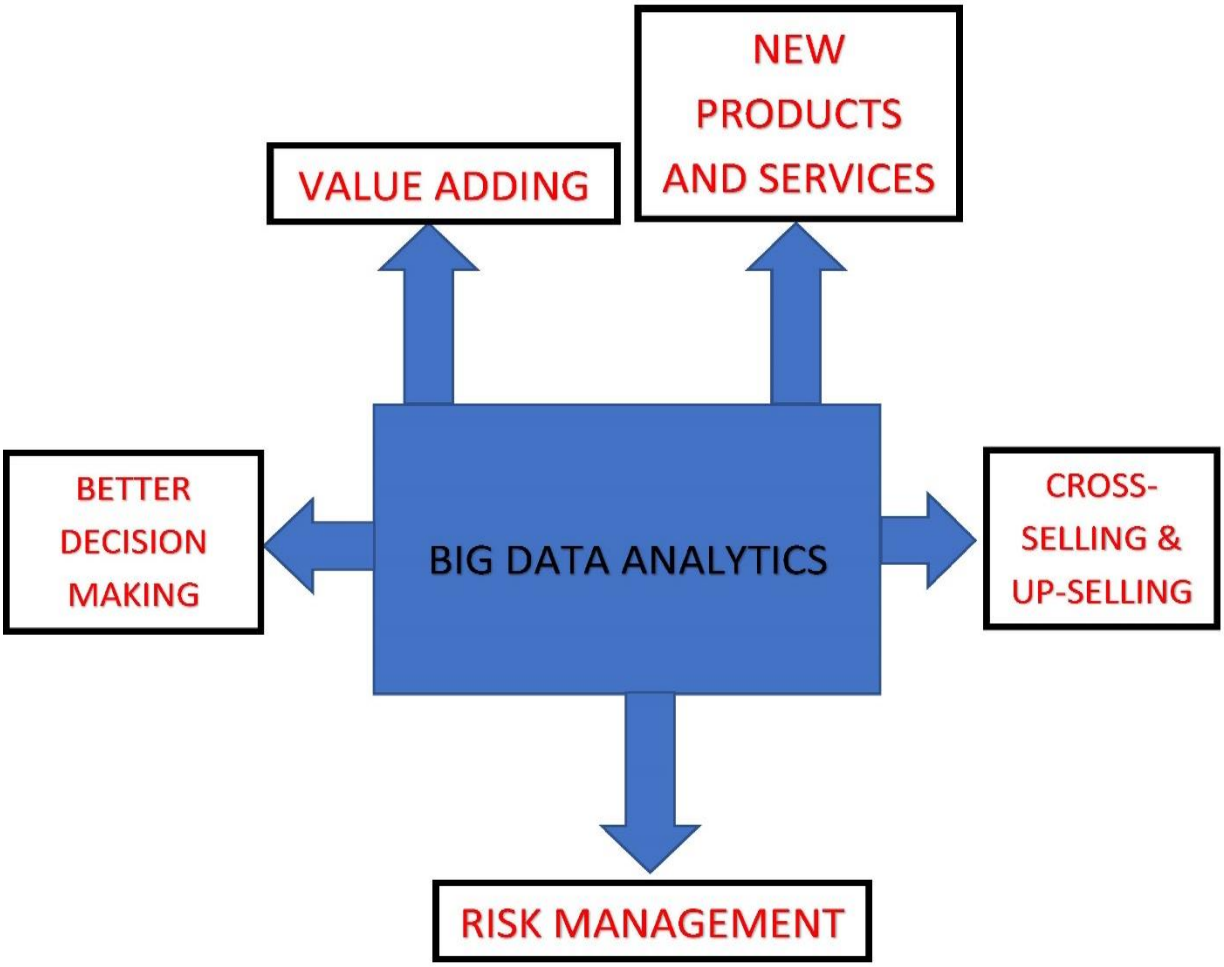


Figure 1.3: Big Data Analytics

1.1.5 BIG DATA CHALLENGES

While big data holds tons of promise, it's not without its challenges. First, big data is…big. Although new technologies are developed for data storage, data volumes are doubling in size about every two years. Organizations still struggle to stay pace with their data and find ways to effectively store it.

But it's not enough to only store the info. Data must be wont to be valuable which depends on curation. Clean data, or data that's relevant to the client and arranged during a way that permits meaningful analysis, requires tons of labor. Data scientists spend a lot of their time curating and preparing data before it can actually be used.

Finally, big data technology is changing at rapidly. a couple of years ago, Apache Hadoop was the favored technology won't to handle big data. Then Apache Spark was introduced in 2014. Today, a mixture of the 2 frameworks appears to be the simplest approach. maintaining with big data technology is an ongoing challenge.

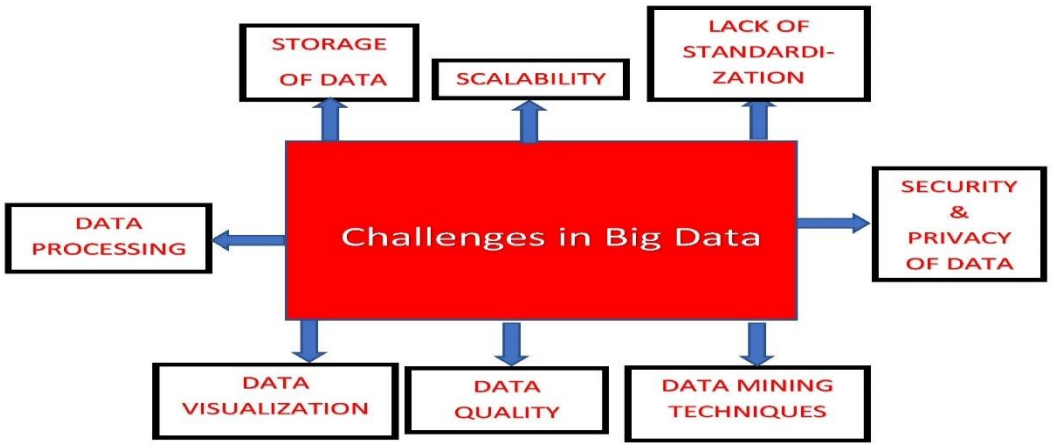


Figure 1.4: Big Data Challenges

1.1.6 HOW BIG DATA WORKS

Big data gives you new insights that open up new opportunities and business models. Getting started involves three key actions:

I. Integrate

Big data brings data together from many sources and applications. Traditional data integration mechanisms, like ETL (extract, transform, and load) generally aren't up to the task. It requires new strategies and technologies to research big data sets at terabyte, or maybe petabyte, scale. During integration, you would like to usher in the info, process it, and confirm it's formatted and available during a form that your business analysts can start with.

II. Manage

Big data requires storage. Your storage solution is often within the cloud, on premises, or both. you'll store your data in any form you would like and convey your required processing requirements and necessary process engines to those data sets on an on-demand basis. many of us choose their storage solution consistent with where their data is currently residing. The cloud is gradually gaining popularity because it supports your current compute requirements and enables you to spin up resources as required.

III. Analyze

Your investment in big data pays off once you analyze and act on your data. Get new clarity with a visible analysis of your varied data sets. Explore the info further to form new discoveries. Share your findings with others. Build data models with machine learning and AI. Put your data to figure.

1.1.7 BIG DATA TECHNOLOGIES

I. Apache Hadoop

Apache Hadoop may be a java based free software framework which will effectively store great deal of knowledge during a cluster. This framework runs in parallel on a cluster and has a capability to permit us to process data across all nodes. Hadoop Distributed File System (HDFS) is the storage system of Hadoop which splits big data and distribute across many nodes in a cluster. This also replicates data during a cluster thus providing high availability.

II. Microsoft HDInsight

It is an enormous Data solution from Microsoft powered by Apache Hadoop which is out there as a service within the cloud. HDInsight uses Windows Azure Blob storage because the default filing system. This also provides high availability with low cost.

III. Hive

This is a distributed data management for Hadoop. This supports SQL-like query option HiveQL (HSQL) to access big data. This can be primarily used for data processing purpose. This runs on top of Hadoop.

IV. Sqoop

This is a tool that connects Hadoop with various relational databases to transfer data. This can be effectively wont to transfer structured data to Hadoop or Hive.

V. Polybase

This works on top of SQL Server 2012 Parallel Data Warehouse (PDW) and is used to access data stored in PDW. PDW is a Datawarehouse appliance built for processing any volume of relational data and provides an integration with Hadoop allowing us to access non-relational data as well.

VI. Big Data in Excel

As many of us are comfortable in doing analysis in EXCEL, a well-liked tool from Microsoft, you'll also connect data stored in Hadoop using EXCEL 2013. Hortonworks, which is primarily working in providing Enterprise Apache Hadoop, provides a choice to access big data stored in their Hadoop platform using EXCEL 2013. You can use Power View feature of EXCEL 2013 to easily summarize the data. Similarly, Microsoft's HDInsight allows us to connect to Big data stored in Azure cloud using a power query option.

VII. Presto

Presto developed by Facebook and Facebook has recently open-sourced it's Query engine, made to handle petabytes of knowledge. Unlike Hive, Presto doesn't depend upon MapReduce technique and may quickly retrieve data.

1.2 INTRODUCTION OF SMART CITY

A smart city is basically one that utilizes innovation to improve results over each part of city activity and upgrade the administrations it offers to its inhabitants. There are a wide range of motivations to concentrate on making our urban areas savvy, over half total populace is living in urban communities and there are 60% urban communities around the world that are yet to be

assembled. Urban areas are at the vanguard of worldwide development and by 2050, over 70% of our total populace will live in the urban areas. While looking at making our traditional urban communities keen, we ought to likewise consider that shrewd urban communities ought to be resident driven, collective, responsive, practical, responsible and straightforward so as to help individuals in a most useful way.

1.2.1 BIG DATA'S ROLE IN THE SMART CITIES

Let us envision a presentation on your dashboard window alarms you that climate conditions have made your standard course to work less positive and, hence, reroutes your excursion to the workplace dependent on ongoing computations of ideal conditions. Once inside the parking structure, another alarm informs you of the nearest parking space, controlled by an estimation of your work environment dependent on your set up driving examples.

Such a dream isn't one of the prospects yet of present reality. Surely, urban areas have progressively embraced advances like Big Data, the Internet of Things (IoT), and disseminated sensors to make what many are calling the city of things to come. We can see it through the arrangement of network assets (network fiber, civil Wi-Fi) or single reason applications (keen road lighting, brilliant stopping or waste administration). A couple of urban areas have started to esteem an increasingly forward-looking methodology that cuts over different Smart City applications, concentrating on information as the normal component [11].

Information is the soul of a keen city. Since information is the new gold, so by utilizing sensors or by deriving information from various devices like mobile phones etc., large database of information can be gathered and Big Data investigation can be utilized on this data to manage the smart structures of a smart city [12]. The smart city enormous information system must ensure an effective stockpiling of the shifted information structures (organized, to un-organized and semi-

organized). They should have the ability to process both continuous just as recorded information (various necessities of ongoing versus clump handling). In addition, they should give adaptability regarding information stockpiling and handling (in case of an unexpected increment in load). At long last, they ought to likewise have the option to share the handled outcomes over an assortment of uses/administrations in a gradual and versatile way.

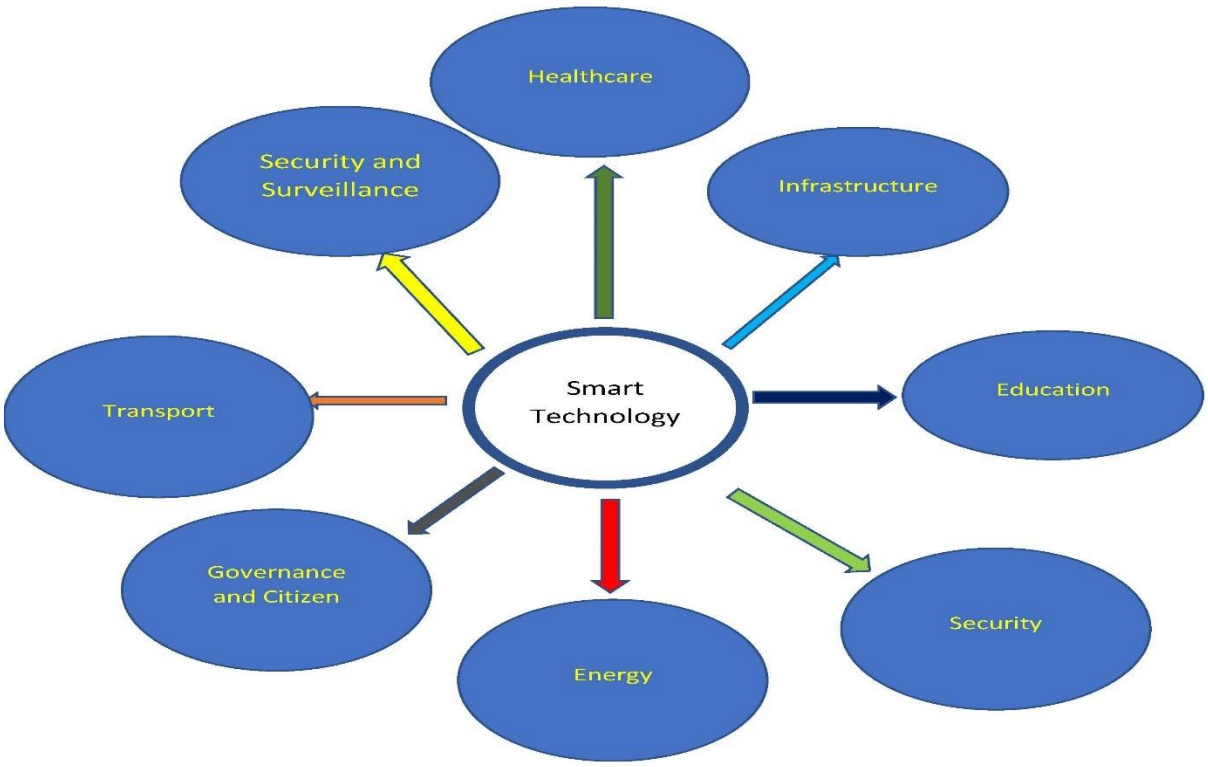


Figure 1.5: Components of The Smart City

1.2.2 SMART CITY APPLICATIONS OF BIG DATA

The utilization of big data (enormous information) propels for the keen(smart) city enables compelling data amassing and taking care of to make information that can update various brilliant city organizations. Likewise, huge information enables leaders to get ready for any extension in brilliant (smart) city administrations and assets. For big data to accomplish its objectives and advance administrations in savvy urban communities, it needs the correct instruments and techniques for productive and viable information investigation.

1.2.2.1 SMART GRID

The fast dissemination of smart grids has empowered scientists to incorporate, examine, and utilize ongoing force age and utilization information, just as different sorts of ecological information [15]. The improvement in vitality effectiveness and wise administrations is relied upon to bring about a high speculation proficiency of the current smart grid framework. For smart grid condition, multitude of data is collected from multiple points of information. The effective utilization of large information gathered from the savvy framework condition can help leaders think of an insightful choice as far as the gracefully level of power while satisfying the requests of the client. The examination of the keen network information can likewise help anticipate the need of the force flexibly later on.

1.2.2.2 SMART HEALTHCARE

During last ten years, a huge measure of information has been produced in the healthcare sector. Legitimate analytics instruments can allow healthcare specialists to gather and analyze patients' information. In addition, legitimate analytics of huge healthcare information can help anticipate

plagues, fixes, and maladies. Moreover, study of health-related data can help specialist track the disease and recuperation status of their subjects which will save lives**.**

1.2.2.3 SMART TRANSPORTATION

The information produced by transport frameworks can help enhance cargo developments **(Docherty et al., 2018**). In addition, the enormous information gathered from shrewd vehicle frameworks can help solidify shipments and upgrade dispatching developments by lessening gracefully chain wastage. Smart transport information can likewise give numerous advantages, for example, decreasing the fuel consumption with the use of analysis of route related data. A data analytics-based model for tackling several issues relating to vehicular moments in smart cities. It can help to prevent congestion, traffic jam among other use cases

1.2.2.4 SMART GOVERNANCE

Big Data based investigation can assist governments with setting up and actualize good approaches since they are now acquainted with the requirements of the individuals as far as wellbeing, social consideration, training, etc.

1.2.3 TECHNIQUES & TOOLS

Various methods and ways are proposed for Big data approach application on Smart cities. The three alternate points of views are as follows:

- ✓ Tools and methods utilized for information demonstrating, capacity, management.
- ✓ Tools and techniques involve in scrutinizing the information.
- ✓ Tools, methods utilized for overseeing security and protection.

1.2.3.1 DATA MODELS

Social databases are as yet the most generally utilized capacity instrument for shrewd city applications. Various devices and strategies are utilized to store, move and oversee enormous information in smart urban communities. This is particularly testing a direct result of the idea of enormous information talked about beforehand. Therefore, information stockpiling arrangements, are generally utilized. Because of its continuous updates (speed), huge information is additionally, frequently naturally, time arrangement information, and therefore time arrangement databases, for example, Influx DB are likewise found in writing.

1.2.3.2 COMPUTING MODELS

Similarly, as with information models, the idea of huge information likewise requires diverse registering models. In addition, since savvy urban communities are intended to give open data administrations to residents, this likewise requires exceptional registering models. Cloud, mist and edge figuring are significant and developing advancements in such manner **(Bibri, 2018).** On the backend appropriated information preparing advancements, for example, Hadoop and Microsoft Azure are mainstream. Our overview additionally found that AI is by a long shot the most widely recognized strategy for performing huge information investigation.

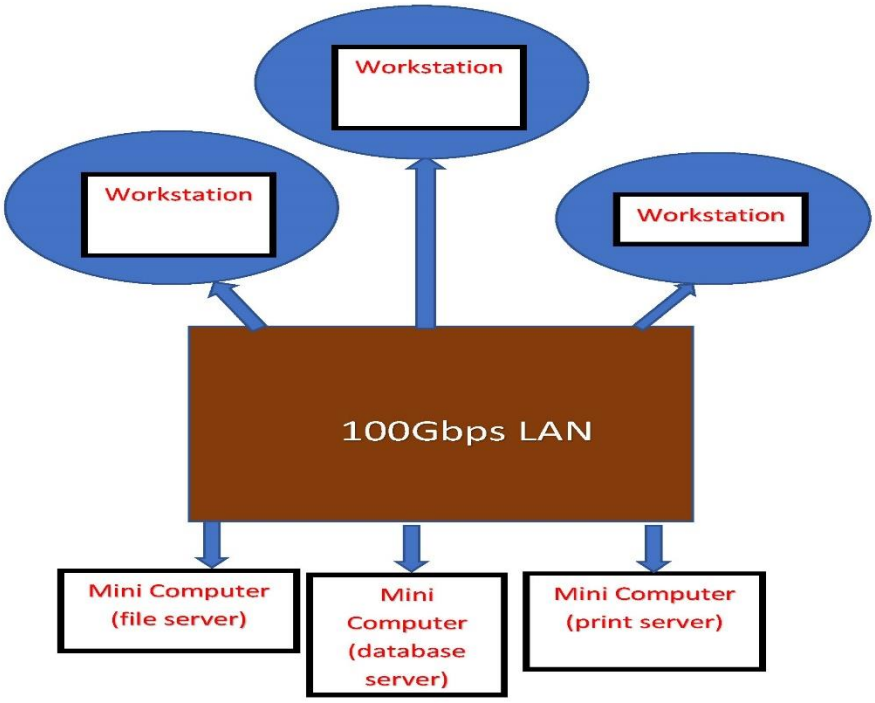


Figure 1.6: Computing Model

1.2.3.3 SECURITY AND PRIVACY

This is a significant zone of keen urban communities is made much progressively significant because of the ongoing focus on security and protection in the IT space by and large. The presentation of General Data Protection Regulation (GDPR) is one consequence of such a core interest. Subsequently, research around there is increase.

1.2.3.4 MARKET DRIVERS

Smart city advances are being driven by significant players, for example, IBM, Deloitte and a lot increasingly modern and private players.

1.2.4 A BETTER CITY THANKS TO DATA

- We should examine how information is helping urban communities around the globe:
- The city of Nanjing, China, has brought sensors into 10,000 taxis, 7,000 vehicles and 1 million private vehicles. The ensuing data is moved step by step to the Nanjing Information Center, where experts can bind together and separate traffic data and send updates to residents on their propelled cells. With these data bits of information, government specialists have made new traffic courses to improve blockage, without consuming money on new lanes.
- Trenitalia, Italy's critical rail director, presented sensors on the trains and now gets persistent declarations on the mechanical condition of each train and upkeep figures that licenses Trenitalia to structure a methodology before a stunning event. These mechanical improvements give adventurers a reliable system and organization, while allowing urban regions to forestall significant disturbances.
- Los Angeles is suppressed streetlights 4,500 miles with the new LEDs. Especially in the more popular streets it will be a mutual structure similar will not, however, new light which results will be educated about the city each bulb conditions. If one breaks, It is perceived and fixed rapidly. Afterward, we can supplant the lights changing tones or flash to caution occupants of various circumstances.
- Large social affairs of people mean immense measures of data is delivered. Huge data is being used to get when, how, and why gatherings structure, and to predict their turns of events and exercises. There are a large number of sensors set up starting at currently,

watching various things. In the near future, these sensors will increment until they can screen everything from streetlights and trashcans to road conditions and essentialness use.

1.2.5 THE DATA CHALLENGE

The compelling administration of information isn't restricted to information catch and capacity, however should likewise incorporate information that is shared and consolidated so it very well may be gotten to, examined and utilized across offices, among associations, and even with the network on the loose. In each significant city in the United States, and past, there are a great many sensors creating a stunning measure of information each millisecond, second, moment, hour, and day. That information is caught, put away and pretty much overlooked after. Keen urban areas should be based on systems that take into account the free correspondence of information.

Information sharing speaks to both a necessity and an open door for Smart City arrangements. Unmistakably information sharing across citywide divisions and stages is a fundamental component of any Smart Cities plan. Our expectation is that most urban communities will execute information sharing as a component of a developmental excursion from information incorporation to information trades, and afterward to information commercial centers and the big data analytics challenges and smart cities solutions mentioned through the table. 1 given below.

Area Effected by Big Data	**Data related Smart City Issues / Major Challenges in Data Models**	**Possible Workout Researchers**

Effects Related to Environment	--Collecting and putting away information through sensing elements --Occasional truancy of data	--Store information in conveyed and time arrangement databases --Develop models that enhance equal I/O.
Effects Related to Economy	--Health care difficulties --AI approach for Economic analysis	--Information Collection from Danish males 30 to 60 age group **(Druedah and Munk-Nielsen, 2018)** --Collecting conduct information
Effects Related to Transportation	--Dissecting information from various sources --Counting the vehicles passing by **(Chauhan, S., 2016)** --Big Data Analytics to diminish the Congestion **(Abberley et al., 2017)**	--Semantics from different sources and for different kinds of information.

		--Ontology Definitions

Table 1.1: Overview of Big Data Analytics Challenges and Solutions in Smart Cities

1.3 RESEARCH OBJECTIVES

I. Analysis and evaluation of different analytical approaches in order to lay foundation for data integration in smart transportation in smart cities.

II. Analysis of challenges that are associated with smart transportation and data associated with it. The analysis will also shed light upon best practices to minimize such challenges and make integration more flexible and modular.

III. To presents a data analytics-based model for tackling several issues relating to vehicular moments in smart cities. The model can help to prevent congestion, traffic jam among other use cases.

1.4 RESEARCH GAP

The proliferation of big data provides numerous opportunities for data-based smart city transformation. However, as with any large-scale initiative for change, the move to data-based smart city transformation is not easy. The challenges in executing smart city projects that depend on using urban big data need to be determined. The challenges that arise during the

transformation of data to information for smart cities are data quality management, integration of data from different sources, data privacy concerns. The aim of this research is to minimize such challenges and provide a comprehensive overview and model of integrating big data in smart transportation of smart cities.

1.5 THESIS ORGANIZATION

In Chapter- 2, smart transportation is discussed. In Chapter 3, the literature related to the problem is discussed. In Chapter 4, Model is presented. Optimum solutions and numerical examples are presented as well. In Chapter 5, presents result analysis and discussion. Chapter 6, presents conclusions and suggestions made for future research.

1.6 SUMMARY

The big data problem arising from the construction of a smart city is not only a cutting-edge problem for next-generation scientific research but also a motive that drives the development of a smart city. It brings new opportunities and challenges and urges us to expedite technology innovation and research related to big data, thereby promoting and accelerating the development of the intelligence service industry, better implementing the various intelligence applications in a smart city, allowing for more and better intelligent applications to serve the entire society, and making city operations more scientific, efficient, low-carbon and safe.

CHAPTER: 2

SMART TRANSPORTATION

2.1 INTRODUCTION

The smart transport revolution is taking the planet by storm. Consumers have more transport options than ever before and completely new ways to get and consume transport services. Organization have access to new, world markets for transport and route optimization benefits. City governments have new opportunities to incentivize green, efficient modes of transport, to reduce congestion and improve the quality of life.

2.2 WHAT IS AN INTELLIGENT TRANSPORTATION SYSTEM (ITS)?

An Intelligent Transportation Systems (ITS) is an efficient transportation and mobility system utilized in smart cities. It takes advantage of technologies like the web of Things (IoT) and large data analytics to manage traffic and mobility, enhance transport infrastructure, and supply improved interfaces for transport services. The ITS market is growing, expected to reach $40 billion worldwide by 2024, according to Grand View Research.

2.2.1 BENEFITS OF AN ITS

* Making interconnected transport systems with openly communication between devices and vehicles.

* Actively managing traffic, helping conveyance to stay on schedule.

* To ensure peoples have access to real-time information about traffic and public transport conditions.

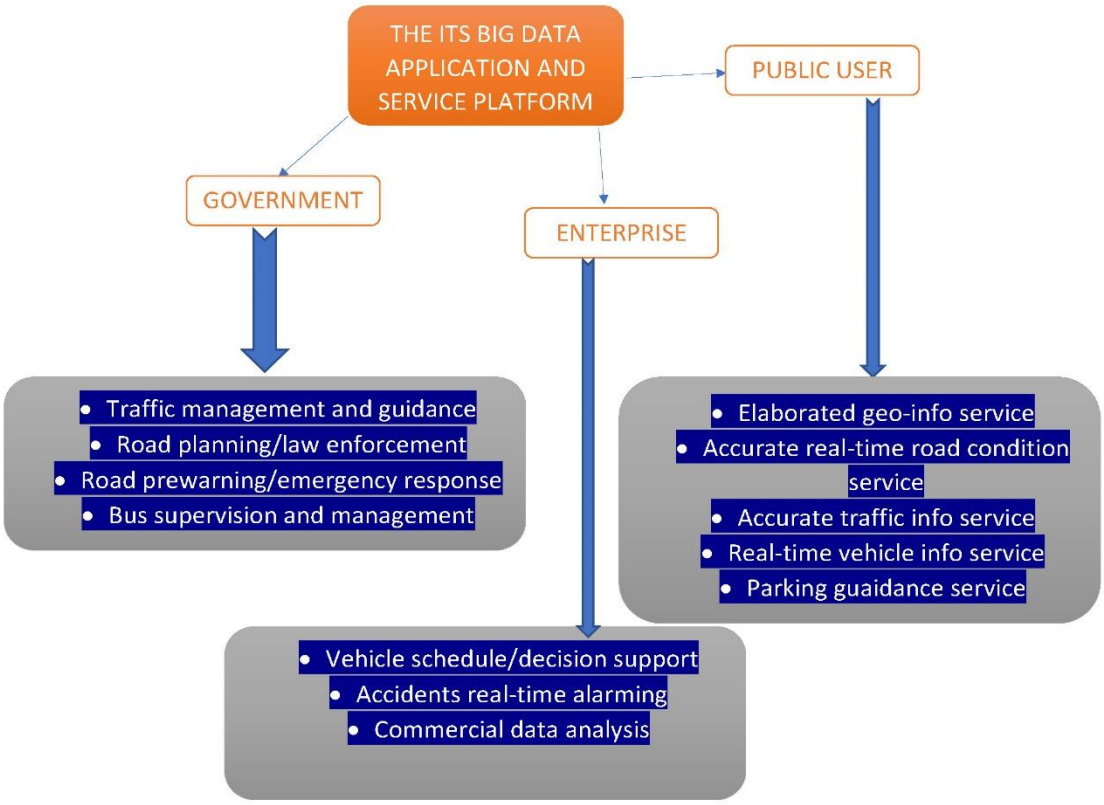


Figure 2.1: ITS Big Data Application

2.2.2 HOW ITS WORKS

An ITS has four main components:

I. Traffic Data Collection: The system uses devices such as GPS devices, road cameras, and vehicle identifiers to gather information in real-time. They collect information about the location and speed of vehicles, and traffic conditions.

II. Data Transmission: ITS transmits the data collected by the sensors to a management center where is analyzed and forwarded to applications.

III. Traffic Data Analysis: At this stage, the cleaned data is customized for further analysis, then sent to end-user interfaces.

IV. Traveler Information: Finally, the information is made available to citizens, who can access it via the radio, web browsers, or text messages. These data inform peoples about traffic and helps them to plan optimal routes.

2.3 HOW IoT IS CHANGING THE FUTURE OF TRANSPORTATION

The Internet of Things is changing the transportation industry by integrating technical and business mobility trends, data analytics, and automation. IoT connects objects via sensors and other devices that collect and send data about real-time network activity.

2.3.1 IoT USE CASES

The goal of every transportation solution varies consistent with the utilization case. For instance, transportation system systems aim to enhance rider experience and safety. Other samples of solutions serving the transportation sector today include:

I. Connected Cars: Cars today believe connectivity, with each vehicle use the web with over 50 microcomputers to watch the brakes and engine, control the tire pressure and exhaust gas composition. New technologies like Vehicle-to-Infrastructure (V2I) and Vehicle-to-Vehicle (V2V) allow users to order parking spots, prevent collisions, and promote smooth traffic flow.

II. Vehicle Tracking System: Typically employed by fleet operators for tasks like routing, fleet management, dispatching, on-board information, and security. The IoT has improved their performance including the power to capture driver behavior to attenuate idling and stopping time.

III. Conveyance Management: IoT technology is usually utilized in transportation system management systems. It offers benefits like real-time vehicle tracking, factoring for unexpected events (bad weather, road closure), and supply personalized travel information.

IV. Traffic Management: Smart traffic management relies upon parking sensors, smart traffic signals as a complement to the normal video surveillance, to beat traffic jam. To learn more, see the complete article, IoT in Transportation: Benefits, Challenges, and Uses.

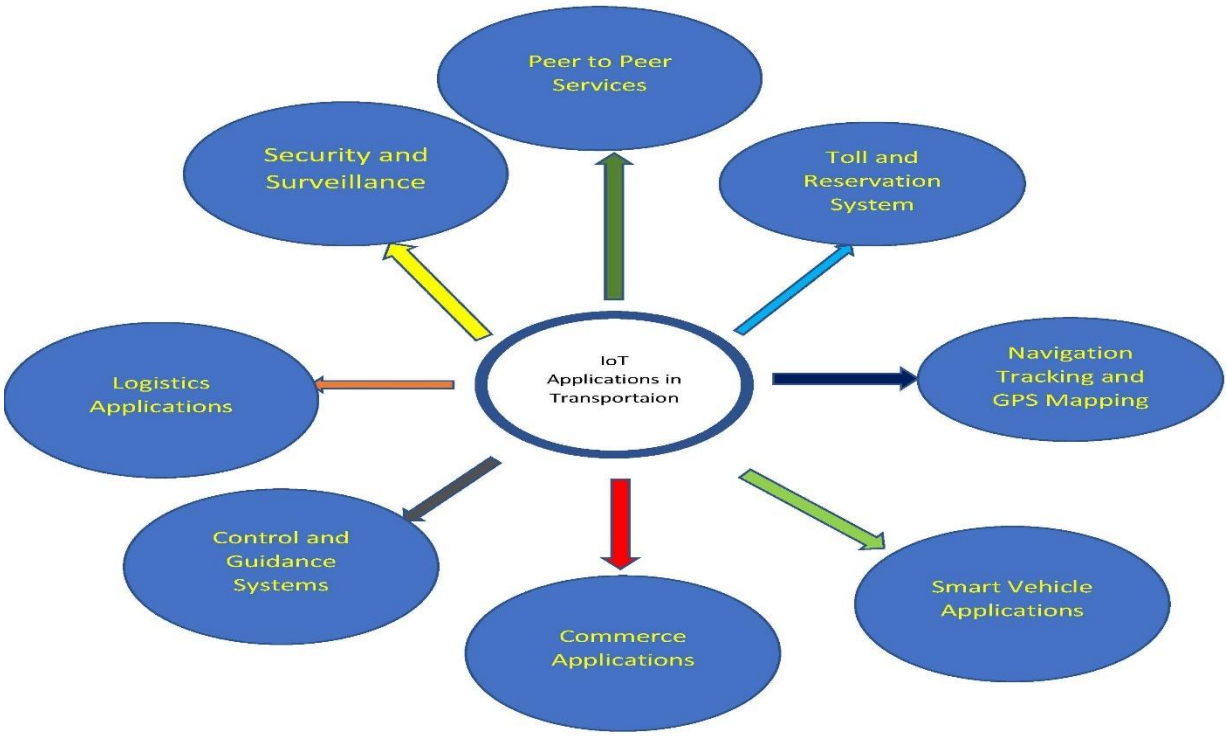


Figure 2.2: IoT Application in Transportation

2.4 ON-DEMAND TRANSPORT

The on-demand economy may be a new paradigm during which economic entities provide on-demand services to others. The modern economy is fueled by a demand for instant gratification. It aims to supply an instantaneous response to the stress of the buyer by bringing opposing parties within the supply chain closer together and building transport infrastructure round the customer experience. Thanks to the web, consumer behavior has changed, with convenience becoming a critical influence on purchase decisions. A central a part of the on-demand economy is on-demand transport services.

2.4.1 WHAT IS ON-DEMAND TRANSPORT?

Consumers in the digital era demand prompt service. Online order for cabs, car hires and other modes of transportation are the easy nowadays. On-demand transport applications can process payments quickly and provides peoples access to a car or bike in minutes, using their mobile phones.

Some of The Trends Becoming Popular Are:

I. Ride-Hailing: a web platform connects passengers in need of a ride with nearby drivers heading an equivalent way. Rides are booked via an application.

II. Car Sharing: Private car owners rent their car to people through an app, allowing any visitor or resident book a car for a couple of hours.

III. Ridesharing: people that seek one-way rides can find available seats offered by drivers getting into an equivalent direction.

IV. Carpooling: Rides are exchanged between family and friends who alternate using their cars to move the group. Employees use this technique to share rides to figure also as parents taking turns to move their children.

2.5 SMART CITY TRANSPORT

A smart transportation can make an enormous difference to the way passengers commute in dense urban areas, saving costs for municipalities, and providing better and safer service to city residents.

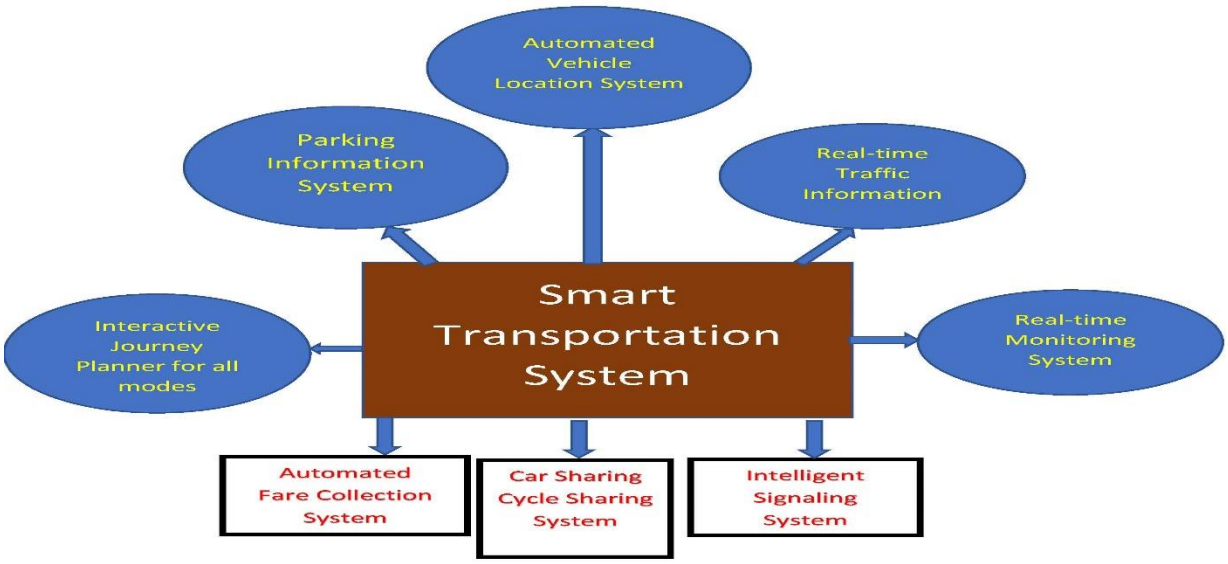


Figure 2.3: Smart Transportation System

2.5.1 WHAT IS SMART CITY TRANSPORT?

Smart city transport is that the use of electronic, wireless, and Internet technologies to supply access to smarter, safer, and faster travel between two points during a large city. It provides city authorities with information and control the traffic flows. Cities who wish to become a sensible city often start by deploying a sensible transport infrastructure, within the sort of an Intelligent Transport Network (ITN).

An ITN Includes:

* A Public Transportation Management system.

* A route information system and electronic timetable.

* A safety and vehicle control system.

* A single fare cards.

2.5.2 BENEFITS OF SMART TRANSPORT

Some of the advantages of smart transport for cities and communities include:

I. Improved Quality of Life: More accessible and efficient public transportation reduces expenses and improves the standard of life for city residents.

II. Reduced Pollution: Smart transport relies heavily on environmental policies, like promoting the utilization of conveyance. This reduces private car usage, reducing emissions.

III. Improved Conveyance Safety: Monitoring and tracking of the general public transportation network can help respond quickly to incidents and emergencies.

IV. Mobility Marketplace: the rise in passenger traffic creates a requirement for mobile transportation apps, which help people consume transportation services across the town .

V. Smart Parking Solutions: With the help of assistance of sensors, security cameras, and Internet connectivity, cities can reduce the problem of parking in congested areas by sharing data about available parking to peoples via mobile apps.

2.6 INTEGRATED MOBILITY

City populations are growing exponentially, aggravating urban challenges like congestion and pollution. Integrated mobility may be a systematic approach to solving these issues, combining physical infrastructure and town planning with technology. This approach leverages intermodal transportation, allowing passengers to mix several modes of transport. for instance, a passenger might want to rent a motorcycle to a railway station, pass by train to a different city, then order a taxi to urge to their final destination.

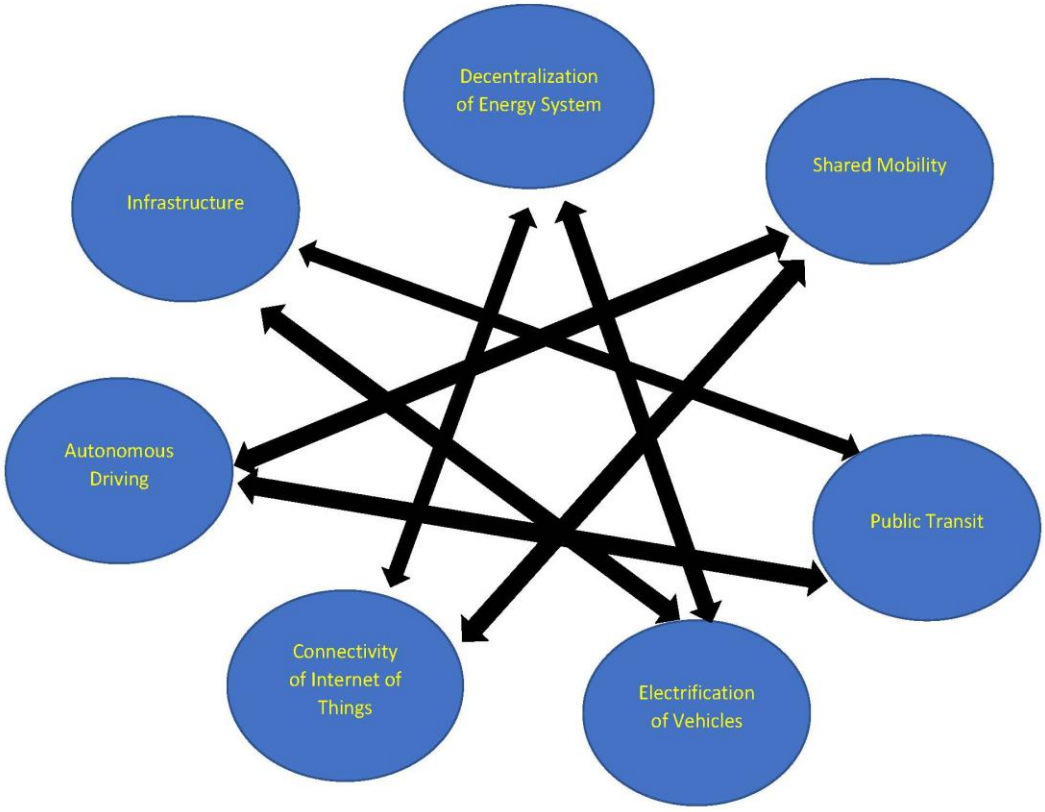


Figure 2.4: Integrated Mobility

2.6.1 INTEGRATED MOBILITY BENEFITS FOR SMART CITIES

I. Economic: It providing people with better access to their activities.

II. Environmental: Efficiently using public space and reducing car usage to supply more green spaces and reduce pollution.

III. Reducing traffic congestion: Promoting conveyance reduces the necessity for personal vehicles.

IV. Commuting Flexibility: Passengers can use cloud platforms and apps to arrange their end-to-end journeys, book trips, and vehicles in real-time, paying electronically.

2.7 SMART TRANSPORTATION TRENDS TRANSFORMING TODAY'S VEHICLES

Given the exponential growth of eCommerce, the demand for shipping and efficient transport methods is on the increase. Many studies show that shipping may be a prime concern for consumers, and improved shipping options have a serious influence on online purchasing decisions.

However, the transport industry faces several challenges. Another issue **is** that the shortage of warehousing space. to beat these challenges and improve efficiency, organizations are turning to new transportation technologies.

Some of The Trends Transforming Today's Transportation Include:

I. Hybrid Electric Trucks: Besides their electric engine, these trucks implement technologies like brakes that convert K.E. into electricity.

II. Hybrid Electric Buses: These combine the normal fuel engine with an electrical system, which is cost-effective and reduces emissions.

III. Autonomous Vehicles: Several cities are implementing autonomous vehicles as a replacement public transportation option. Such vehicles are already operating in London's Heathrow Airport, and therefore the Charles de Gaulle Airport in Paris.

IV. Delivery Robots: Many startups are launching robotic delivery services, especially within the USA. It gives citizens real-time data about traffic and conveyance.

2.8 APPLICATION OF BIG DATA IN INTELLIGENT TRAFFIC SYSTEM

With the rapid development of society, transportation industry is additionally facing unprecedented challenges:

I. Transportation industry can't meet the rapid growth of data. Transportation industry data's have rich sources, diverse types, and new data is produced. customer information of rail way, road traffic, aviation industry, public transit, are recorded, and tens of billions travel records are generated every year.

II. Operating data generated by transport companies, like the info generated by the courier companies. Dynamic data generated by various sensors, like coil at bayonet point, infrared detector, microwave detector, ultrasonic detector, laser detector, video detector, and so on, and therefore the data are generated by GPS vehicle location tracking system and other mobile device search year the quantity of knowledge generated by the transportation industry during a city has exceeded TB level, are developing from PB level to EB. A massive data space for storing and equipment are required and it must have fault tolerance and stability.

III. Traditional data processing systems are faced with the problem of inefficient or even failure. The information system of transportation industry after development for ten years has had a certain foundation and scale, but generation of new business, rapid growth of data, complexity of the data processing is not foreseen. The traffic information management system using traditional processing technology can't meet the rapid highness of knowledge, this is inefficient, and failures have occurred, when processing big data recently years, new projects

and renovation of the old system were carried out in some places, The rapid growth of data is still not considered, In the course of project construction and maintenance, the development was emphasized, the upkeep was overlooked, data haven't been excavated deeply, with the change of leadership, the life cycle of the system is shortened.

IV. The current management system appeared single function, the lack of integration, backward technology and other issues. In the process of building the transportation data system , homogenization is serious, at an equivalent time, development of data technology in several regions isn't balanced, data acquisition is at different depths in several area, and without uniform standards, the executive department for the project only examine and approve it, but supervision and evaluation is lack. The data in most data system scattered in grassroots enterprises, the functional department just collect the report and ledger on a hard and fast period of time, they didn't achieve connections and data synchronization between systems, and do not know well the data produced by enterprises.

The appearance of massive data for resolving the above questions provides a replacement technical approach, big data applied to transportation industry has the subsequent advantages:

I. Traffic management system used big data technology can handle vast amounts of complex and diverse data.

There are three questions: data storage, data analysis and data management resolved by Big Data. Hadoop system is born with the power to handle massive amounts of knowledge, Data is stored on different nodes. A large task is split into small tasks, and be finished in MapReduce model. At an equivalent time, its stability and fault tolerance are vital. Hive as a

knowledge warehouse can save big data in HDFS, its HQL sentence is translated into MapReduce task, and be executed on different node, HBase as a database can store and operate the info in column mode. Sqoop can translate data between RDMS and Hadoop. Flume may be a highly available, highly reliable systems, it can collect distributed massive logs, and aggregate, transport them.

II. Big data can improve the efficiency of transportation industry. Transportation industry, involving many aspects of labor , got to handle massive amounts of knowledge , has more control model of application, features a batch of kit , if a touch accident occurred, the whole system will run into inefficient state, after using big data technology, the knowledge system can process the info and find out the accident in blast , automatically handle it, or reported to the management staff and ask them to form decisions. Big data has a good predictive ability, it can reduce the probability of false alarm and under reporting of traffic incidents. Traffic guidance is a crucial a part of intelligent transportation systems. By publishing information for travelers, it can show traffic conditions of downstream road, allow travelers to move on the proper travel path, improve the traffic situation within the city. In the aspect of improving transport, improving the capacity of the road network, adjusting traffic on demand, big data technology has very useful advantages.

III. Big data can improve the safety level of traffic. The real-time operation capabilities of massive data can accurately probe traffic accidents, its predictive ability can effectively predict the occurrence of traffic incident, using microwave detection systems, video surveillance systems, mobile detection system, we will build an efficient security model to enhance the security of vehicles. When security incident happened, and emergency rescue needed, due to its

comprehensive processing and decision-making capability, rapid response capability, big data can greatly improve the power of emergency rescue, and reduce casualties and property losses.

2.9 ARCHITECTURE OF INTELLIGENT TRANSPORTATION ON BIG DATA PLATFORM

Intelligent transportation on big data platform may be a combination of multiple systems, models, department, technology. It can be said, it is a comprehensive system of system science, management science, mathematics, economics, behavioral science, and information technology. From the architecture, the platform includes basic business layer, data analysis layer and information publishing layer the basic business layer is the foundation of data analysis layer and information publishing layer, its main function is to finish the essential work of the varied business units, and to supply basic business data. It includes traffic information collection system, signal control systems, video surveillance systems, illegal evidence forensics system, 122 alarm receive and dispose system, GPS vehicle location tracking system, traffic guidance system, vehicle information management system, driver information management system, PGIS system, and so on. the service of basic business layer is that the basis for the varied business units, its data comes from data acquisition system mentioned above, storage and handling of data is very important. Therefore, cloud computing technology are often used on the essential business layer, decentralized system is often integrated into the cloud, this may make sure the security and stability of the application system, and provide an efficient computing environment.

According to the knowledge of the road network, the demand of public travel and comprehensive analysis of knowledge, data analysis layer uses big data technology, data processing work of the

technology, combines with a variety of mathematical models for real-time effective analysis. It can grasp the condition of the transportation systemin any time, such as road congestion degree, average speed, saturation, occupancy rate, interrupt rate. It can make further congestion warning, traffic guidance and other intelligent transportation behavior. Data analysis layer is made on Hadoop ecosystem, use commercial cheap server as hardware platform, use the open-source Linux as OS , use HDFS as filing system for giant data storage, use MapReduce as a parallel computing model, use HBase because the database for processing the info , use Hive as data warehouse, use Sqoop and Flume as tools for data integration.

The information publishing layer according the results of the info analysis layer, publishes traffic conditions to public, business units, industry executives, etc. by internet, mobile terminal, desktop application, report, for his or her travel and business decisions. It is necessary for friendly interface, operating easily, rich feature. The information published include traffic condition, traffic warning, data charts for decision. With the event of the days, publishing channels become diversified, changed from traffic radio and knowledge bulletin board to today's traffic radio, mobile TV, microblog, WeChat, bulletin board and other forms and channels.

Reference	**ITS privacy and security Issues**	**Solution**
[47]	Re-identification attack by using location trajectory data	K-correlation Privacy Model

[48]	Privacy problem related to the Car Floating Data (FCD)	An architecture to protect the vehicle privacy by combining differential privacy with a 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT) security mechanism. They use a policy-enforcement framework like PRECIOSA PeRA
[49]	Location Privacy Issue	Secure and Efficient Location-Based Service (SELS) scheme that overcome with the privacy issue by making the secret information secure with low computation and communication

		cost
[50]	Location privacy Issue	Autonomous Privacy Preserving Authentication Scheme

Table 2.1: Different type of Security and Privacy Issues and the Proposed Solution in Intelligent Transportation System

CHAPTER: 3

LITERATURE REVIEW

3.1 INTRODUCTION

Welcome to the era of either big data, as we all know this is the era of big data and every model in this era is directly or indirectly by utilizing the concepts of big data. Most important aspect of big data is to handle volume, velocity and Variety. While implementing the concept of smart city these features of big data will play an important role to efficiently manage various services. Here we are trying to find out most important and directly affecting literature. We have been trying to gather literature since 2009.

3.2 LITERATURE REVIEW

The change in standpoint from the space of regular work region enrolling to an unyieldingly refined preparing **(Abberley, L et al., 2017) [2]** similarly as the basic addition in related contraptions and sensors have made feasible the vision for living in an astute circumstance**,** on account of the development of urban populace and quick urbanization. Presently, urban execution relies upon the physical structure just as on the openness and nature of data correspondence and social establishment **(Aguilera, U et al., 2017) [3]**. The key engaging specialist of these sharp city applications is possibly the Internet of things wherein regular items and gadgets are associated with the system advances. The guarantee of a savvy or smart city prompts an exponential increment in information by a few significant degrees. Thus, such colossal volumes of information or huge information are at the center of the administrations rendered by the IoT **(Al Nuaimi et. al., 2015) [4]**. Enormous information Harmful to obtain important information from went information gathered from different resources makes the city able. The most part such characteristics of the information included went disorganized Highlights and vast information collected by different methods. The location of the smart advances with huge data and appropriated processing, in which

diverse sharp applications exchange information using embedded sensor devices and various contraptions facilitated with the distributed computing foundation to create a lot of unstructured information [7, 13]. These a lot of unstructured information are gathered and put away in the cloud or server farm utilizing circulated shortcoming open minded databases, for example, not just SQL, which is used to improve a lone help or application and is shared among different administrations [3]. In this manner, the programming model for handling enormous informational indexes with equal calculations can be utilized for information investigation to get an incentive from the put away information **(Anisetti, M. et al, 2018) [5]**.

The significance of huge information investigation utilizing AI, information mining, AI or factual strategies in savvy urban areas can't be focused on. We led a broad writing survey of the different manners by which huge information investigation are being applied to smart urban areas. In the field of brilliant urban areas huge information investigation is the key innovation to empower proof based dynamic, creating and offering new open types of assistance just as permitting residents to co-make new open administrations **(Bassoo. et al. 2018) [9].** Huge information can likewise be utilized to screen government activities just as arrangement of different client related open administrations [2]. Checking and anticipating air quality is another case of how enormous information can be utilized in keen urban communities on which there is an abundance of writing **Antonic, A., et al., 2016) [6].** Customary AI draws near, for instance, Support Vector Machines, Decision Tree, Nearest Neighbor classifier are still commonly used techniques for the assessment of enormous data. For gigantic datasets, Support Vector Machine is an ongoing circulated usage of portion bolster vector utilizing subspace apportioning. Bayesian methodology is very mainstream for certain utilizations of keen city administrations utilize Bayesian induction which is an information driven way to deal with distinguish the least vitality proficient structures of an

area with huge populace. **Bokhari, S., & Saiz, A. (2018) [8]** have contended that the greater part of the information created by huge information is unlabeled.

Some More Findings Are Below:

Pradeep Pai T & Shashikala K L, [2019 IEEE International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con)] **[33]** have proposed a paper on **"Smart City Services - Challenges and Approach"**

This paper describes how smart services can be incorporated in city management. Paper describes a micro services-based architecture in realizing these smart services and the implicit benefits of using such an approach. The paper also describes security and privacy concerns in building smart city services and also describes methodologies that can be adopted to mitigate security and privacy concerns.

Lasse Berntzen & Marius Rohde Johannessen & Rania El-Gazzar, (ResearchGate ICDS 2018: The Twelfth International Conference on Digital Society and eGovernments) **[34]** have proposed a paper on **"Smart Cities, Big Data and Smart Decision- making Understanding "Big Data" in Smart City Applications"**

This paper proposed several application areas of smart cities, and related data sources used for decision-making. In many cases, systems may make decisions on their own. Such systems can play an important role in the development of smart cities. In other cases, the data can be combined with historical data or other open data sources play a role as the foundation for decision-making. Our findings are

presented as an analytical framework, which will be used for further empirical studies into this domain.

Hamid Fekri Azgomi & Mo Jamshidi, (2018 IEEE 30th International Conference on Tools with Artificial Intelligence) **[35]** have proposed a paper on **"A Brief survey on Smart Community and Smart Transportation"**

In this survey paper, an attempt made to go through the concept of smart city and smart community. Its different aspects were reviewed and some of their challenges presented. In addition, it indicated that among all categories including Environment, Healthcare, Energy and smart transportation, the latter one has received more attention in recent years. Some recent work on smart traffic control and autonomous vehicles were presented and finally complex path planning framework was discussed as our future challenge for smart cities. At the end, we tried to introduce some of the future challenges in this concept.

Sanjeev Kumar & Anand Prakash, [2017 International Journal of Science and Research (IJSR)] **[36]** have proposed a paper on **"Role of Big Data and Analytics in Smart Cities"**

The aim of this paper is to study the real potential of using Big Data Analytics in Smart Cities. In this work, we studied cases across the globe where decision maker is using Big Data Analytics as a tool for making Smart City. The paper covers how Internet of Things, Machine to machine, Big Data and Smart Cities Linkages can help in doing predictive analytics which can be helpful to human wellbeing.

Hashem, et al. (International Journal of Information Management 36(5) October 2016) **[37]** have proposed a paper on **"The role of big data in smart city"**

In this survey paper, the significant increase in connected devices in urban cities has led to the rapid growth of data, which has elicited the attention of many researchers in different research domains. This paper aims to offer a comprehensive view of the role of big data in a smart city. In this context, we discussed the enabling technologies used in the smart city. The future business model and architecture with the aim of managing big data for smart city were also proposed, and the applications of the smart cities in which big data analytics can play an important role were discussed. Different case studies were also outlined.

Cheng, et al. (2015 IEEE International Congress on Big Data.) **[38]** have proposed a paper on **"Building a Big Data Platform for Smart Cities: Experience and Lessons from Santander"**

In this paper the author presented the architecture overview of a working big data platform called CiDAP and discuss its major design issues. Since October 2014, the CiDAP platform has been deployed and integrated with Smart Santander, one of the largest smart city testbeds in Europe, to collect city data and then serve results to a few real applications.

Kitchin, R. [2014 Cambridge Journal of Regions, Economy and Society, 8(1), 131–136.] **[39]** have proposed a paper on "**Making sense of smart cities: addressing present shortcomings."**

The focal emphasis of this research is to present certain shortcomings that are associated with smart cities. These shortcomings can hamper growth towards being fully autonomous working cities.

Sumedha Chauhan & Neetima Agarwal & Arpan Kumar Kar (2016 info, Vol. 18 Iss 4 pp.) **[40]** have proposed a paper on **"Addressing Big Data Challenges in Smart Cities: A Systematic Literature Review"**

The key objective of this research was to discover the approaches to deal with challenges associated with big data in smart cities. In order to accomplish this objective, a systematic literature review was conducted to identify such challenges and solutions to deal with them as proposed by various researchers. This research papers follows a rigorous methodology for selecting papers for systematic literature review. The selected papers were reviewed, analyzed, and interpreted to identify the potential strategies to deal with big data challenges. The results of this paper contribute valuable information to the big data practitioners by illustrating the process of effective management of big data in smart cities. The process of managing big data in smart cities can facilitate in utilizing the benefits that can be drawn by analyzing the big data in real time.

Gang Zeng (2015 IEEE), **[41]** has proposed a paper on **"Application of Big Data in Intelligent Traffic System"**

This paper proposed discuss the challenges faced by the transportation industry, and advantages of big data used in the transportation industry. Then propose an architecture of intelligent transportation system on big data platform, at last, he discuss the key technology in ITS. of course, ITS should have more advantages in the future.

Zhao-xia & Ming-hua (2019 International Conference on Intelligent Transportation, Big Data and Smart City), **[42]** have proposed a paper on **"A Dynamic Prediction Model of real-time link travel time based on Traffic Big Data "**

In this paper, a dynamic prediction algorithm is proposed based on big data analysis for real-time link travel time. According to the principle component feature extraction of traffic big data in the traffic information platform, the real-time road condition monitoring and travel time prediction are carried out and the real-time road travel time dynamic prediction is carried out on the traffic information platform based on traffic big data analysis results.

Vidya & Deepa (2019 IJRTE), **[43]** have proposed a paper on **"Big Data Analytics in Intelligent Transportation Systems using Hadoop"**

In this paper, the inevitable eventual outcomes of this synthesis consider have emitted an impression of being Intelligent Transportation System is a wide field which covers unmistakable improvements and they perceive a basic improvement in the advancement time run. ITS plans have the conceivable to offer the running with benefits: enhanced security, ampleness, conservativeness, straightforwardness, and multi-reason affiliations. Through the Intelligent transportation structure, unprecedented spaces take tendencies. This paper shows a wide-running territory of sharp transportation framework and its applications and dimension of redesigns. This paper empowers specialists to commonality with Intelligent Transportation structure chart and gives experts data on ITS districts in which further examination might be required.

Awaysheh, et al. (2019 IEEE), **[44]** have proposed a paper on **"Big Data Security Frameworks meet the Intelligent Transportation Systems Trust Challenges"**

Currently, from various considerations, the vendors of ITS appliances prefer to develop proprietary security that reflects their interests. Various communication protocols and standards are typically deployed within each platform. Hence, interconnecting heterogeneous security services provided by different vendors, and providing seamless interactions/interoperations across the available platforms remain a challenging task. A high-level discussion has been proposed to address this issue and solve better the heterogeneity in the security layered. Also, it aids to provide the interoperability for intelligent transportation devices and services from different vendors of technology independent reference architecture. This study concentrated on security classifications for modern ITS development based on analysis of published big data frameworks and use cases. The work aims to facilitate architecture design and technology selection in the construction of secure ITS applications. As so, a reference architecture was proposed, which consists of functional components of the Apache Hadoop security stack. Also, we implement a mapping between the specification domain requirements and the related frameworks and use cases. Finally, an adapted model that presents a horizontal level of multi-layer security service that realizes a security mesh through ITS was presented.

Alsaffar, et al. (2018 3ICT), **[45]** have proposed a paper on **"Smart Transportation System: A Review of Security and Privacy Issues"**

In this paper, one of the trend technologies that is contributed in developing of ITS is Internet of Thing (IoT). It is like any system using Wireless in IoT that can exposed to different types of security and privacy issues. In this paper, the ITS and its components with the focus on Location

Based Services (LBS) was discussed. Furthermore, the different security and privacy issues and solutions proposed by different researchers were discussed including the re- identification attack, privacy issues related to the FCD and location privacy issues. As a future work, a model of privacy and security challenges accompany with solutions for LBS will be proposed based on further research review.

Liu Yang (2018 International Conference on Intelligent Transportation, Big Data and Smart City), **[46]** has proposed a paper on **"Big Data Technology and Its Analysis of Application in Urban Intelligent Transportation System"**

In this paper, analysis shows that the application of big data technology significantly enriched the practical value of urban intelligent transportation system. In the practical application processes, with the help of cloud computing and clustering mechanism, big data can generate a good amount of data acquisition, mass data retrieval and other functions.

3.3 CONCLUSION

The big data issue emerging from the development of a savvy city isn't just a bleeding edge issue for cutting edge logical examination yet in addition a thought process that drives the improvement of a keen city. It brings new chances and difficulties and urges us to facilitate innovation advancement and examination identified with large information, in this way advancing and quickening the improvement of the insight administration industry, better executing the different knowledge applications in a brilliant city, considering more and better shrewd applications to serve the whole society, and making city tasks increasingly logical, proficient, low-carbon and safe. The critical increment in related devices in urban territories has provoked the speedy improvement of data, which has roused the thought of various investigators in different assessment spaces. This

paper intends to provide a long-term perspective on the work of the great deal of information in a curious city. In this particular case, we examine the powerful innovations are used in a curious city. Distinctive contextual analyses were additionally illustrated. At long last, a few open exploration challenges were disclosed to give the examination bearings to the new analysts in the area. Lastly, we infer that big data can assume a significant job regarding increasing important data and for dynamic purposes.

CHAPTER: 4

PROPOSED WORK

4.1 INTRODUCTION

With global changes, urban communities are getting more organized and denser and so there is a dire need for effective transport to provide them with daily necessities. Henceforth we need greater portability however in more astute manner. Transportation constitutes around 7-13% of Gross Domestic Product of several nations. On individual basis, people spend around 9% times of their life while in transport means. But still, many expensive issues stay unsolved including clog, car crashes, air contamination and environmental change because of fuel utilization and particulate

discharge of the vehicles [24]. Consistent endeavors have been done in the past to take care of these referenced issues however minimal huge achievement is made. As of late, Transportation has become an intriguing issue in Internet of Things (IoT) domain and is looked upon as an effective way to solve the issues related to it [21]. In view of IoT innovation, organizations across different travel and transportation industry areas—including railroads, avionic business, aircrafts, cargo co-ordinations and others—have begun catching the information from practically all sort of frameworks or occasions in general society. With headway of cloud innovation to store information and modern information investigation procedures to separate bits of knowledge from information at fast and precision, associations are settling on better educated choices which was not reachable previously [18]. Along these lines information itself has accomplished key and serious resource significance and information examination instruments and advancements are adding to take care of the issues of blockage, mishaps and environmental change [23].

This piece of the study centers to show the headway in information examination innovation, issues and accessible arrangements. Various papers related to the theme have been studied of which the most appropriate and relevant ones are stated below.

A various approach is there to gather information so as to give better transportation the board [21, 29]. For checking the quantity of gadgets entering or leaving the street, various advances are made and have been discussed in this paper. The paper also talks about Global positioning devices that help in tracking location and movement of vehicles. This will help in lowering traffic congestion rate. Various kinds of sensors are additionally introduced in the paper to gauge CO2. To lessen the clog and street mishaps, a major information diagnostic-based model is introduced by [20] to catch the semantics of street mishaps and blockage utilizing metaphysics. The paper utilizes Big information quantitative data to give subjective data about traffic cosmology utilizing Machine learning procedures. The four stages model attempts to plan a theoretical model of clog in sync 1 to build up a metaphysics for blockage thinking about the effect of street mishaps. In sync 2, clog causing measurements because of mishaps are distinguished. In sync 3, the Big information sources applicable to recognized measurements are distinguished to gather the information for investigation [25, 30]. In the stage 4, the recognized measurements and information from distinguished information sources is investigated to recognize designs which can prompt better choice makings to lessen clog. AI calculation Clustering is utilized to dissect diverse traffic and clog [32].

Another significant target to meet from Smart Transportation is having exceptional Climate. Electric Vehicles are viewed as a significant part to accomplish this goal. An overview on information diagnostic procedures dependent on Big information for decreasing outflows for greener savvy city is given spotlight on keen matrices and electric vehicles [28, 31]. Diverse Hadoop based and electronic information systematic methods are introduced. Battery utilization and charging meters are overseen by Hadoop R measurable bundle and stream preparing utilizing Hadoop Pig contents individually. For online arrangements, choice tree calculations, J48 and M5

calculations are utilized on Weka stage to break down the information gathered from NY Independent System Operators (NYISO).

Literature has additionally talked about the frameworks to deal with the general traffic frameworks. A present an ease Smart Traffic System (STS) is additionally proposed to convey traffic refreshes in a split second while propose a traffic the managing framework that attempts to take care of the traffic issues explicitly in India [19, 27]. A Smart Traffic Management System (STMS) to take care of the issue of traffic the board in India particularly blockage. With this advancement, it would appear that that blockage, street mishaps and traffic managing issues will be understood in future urban communities utilizing IoT and information investigative advances generally [26].

4.2 PROPOSED MODEL

Data model for regular and instant update of traffic on each path in the city. This will be done with the help of choosing the parameters of smart city traffic update system.

- Per Hour Traffic
- Buses in Immobilized Mode
- Obstacle (Tree) On the Road

Figure 4.1: Proposed Smart City Traffic Model

The proposed model is comprised of 4 attributes first three attributes are chosen on the basis of most effective parameters for any traffic related prediction.

Attributes:	**4**
Per Hour Traffic	
Buses in Immobilized Mode	
Obstacle (Tree) on the Road	
Slowness in Traffic (%) (Target)	

Table 4.1: Attribute Table

Table 4.1 shows the attributes that are selected to develop the model. These attributes are Slowness in Traffic occurs due to Obstacles on Road, Busses in Immobilized mode and per hour traffic.

Cross Validation

In this section we will perform K overlay cross approval, cross approval is utilized in any AI models to assess the aptitudes. Here k will be 10 Here we run trait assessment utilizing 10-crease cross-approval. This implies the ADTree calculation is run multiple times, in light of the fact that a five-overlap (- F 5) cross-approval is run multiple times to yield 10 distinctive execution gauges and relating positions.

=== Classifier Model (full training set) ===	
Zero R predicts class value:	10.05185

Table 4.2 Classifier Benchmark

Table 4.2, **Zero R** is a classifier that helps to simply predict the category's majority. It is directly useful for determining a baseline performance. That means a benchmark for rest of classification methods.

=== Cross-Validation ===	
=== Summary ===	
Correlation Coefficient	-0.1402
Mean Absolute Error	3.3377
Root Mean Squared Error	4.357

Table 4.3 Cross Validation of Model

Table 4.3, shows, the coefficient of correlation between -1 and 1, highly correlated if value almost one. A classifier that can't fit the concept in the least will probably get essentially zero correlation (it could be slightly negative thanks to chance effects).

Slowness in Traffic percent					
mean	9.6594	9.1637	17.4812	8.2301	5.2984
std. dev.	2.5033	1.8627	3.18	2.2474	1.1104

Table 4.4: Slowness in Traffic

Table 4.4, the mean of the data provides a better fit to the outcomes than do the fitted function values, according to this particular criterion. Is a measure of the amount of variation or dispersion of a set of standard deviation values. Indicates a low standard deviation that set the value of the mean (expected value is also called) are close, while indicates a high standard deviation values that are scattered over a wide range

=== Model and Evaluation on training set ==	
Clustered Instances	
0	23 (17%)
1	75 (56%)
2	23 (17%)
3	6 (4%)
4	8 (6%)

Table 4.5: Training Set Evaluation

Table 4.5, After finding the clustering Weka classifies the training instances into the clusters consistent in the cluster representation and evaluates the share of instances falling in every and each cluster. within the above clustering produced by k-means shows 17% (23 instances) in cluster 0, 56% (75 instances) in cluster 1, 17% (23 instances) in cluster 2, 4% (6 instances) in cluster 3 and 6% (8 instances) in cluster 4.

Table 4.6: Attribute Selection 10-Fold Cross-Validation (Stratified), Seed:1

Average Merit	Average Rank	Attribute
0.061 +- 0.183	1 +- 0	4 Slowness in Traffic percent
0 +- 0	2 +- 0	3 Tree on the Road
0 +- 0	3 +- 0	2 Immobilized Bus

Table 4.6, At the point when leave one out=False, the dataset is decreased to simply the ascribe to be assessed, in addition to the class characteristic, and a cross-approval is run on the diminished dataset to evaluate the chose exhibition measure.

At the point when leave one out=True, the credit to be assessed is expelled from the full dataset and a cross-approval is run on the decreased dataset to create a presentation gauge X. This presentation gauge X is deducted from the exhibition gauge got by running cross-approval on the full dataset. This yields the assessment score.

CHAPTER: 5

RESULT ANALYSIS AND DICUSSION

5.1 RESULT ANALYSIS

Visulised results after traing and testing the data set of an urban city, 10 cross fold validation has beend used. These all behavioural patters has been generated in unsupervised learing mode.

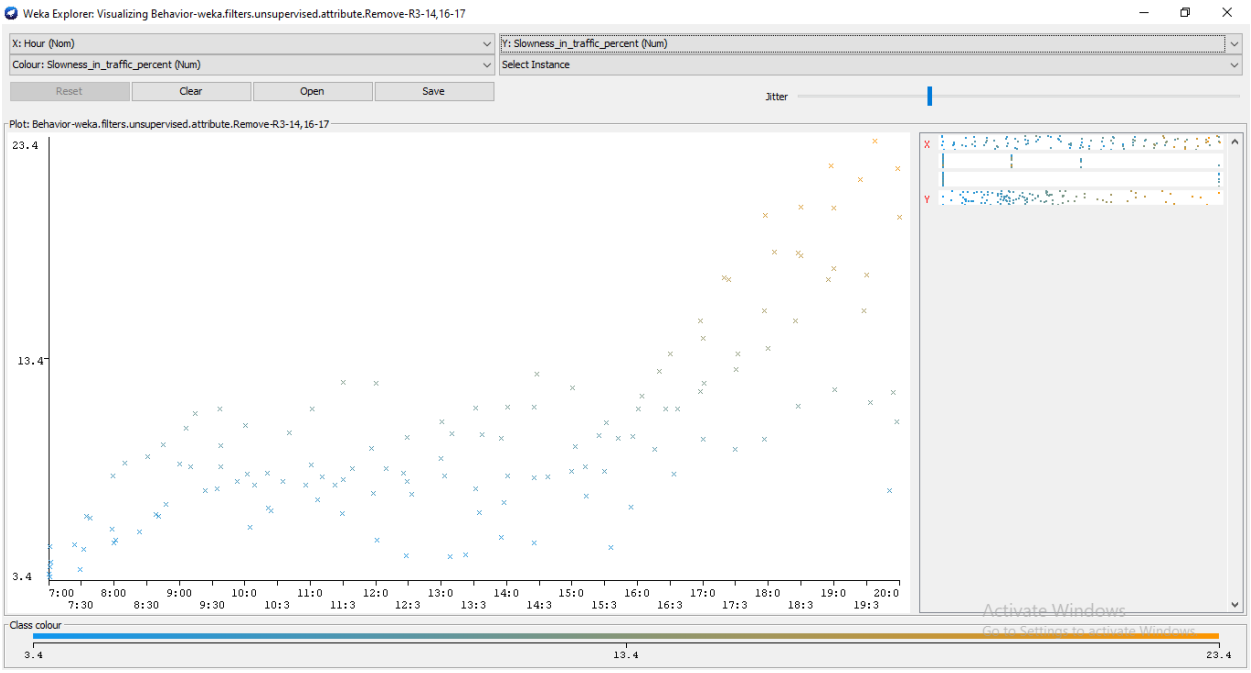


Figure 5.1 Hours vs Slowness

Figure 5.1, is a plot of behavior of traffic in the city based on x axis as slowness in traffic and y axis as hours slot. As on peak hours slot traffic slowness increases

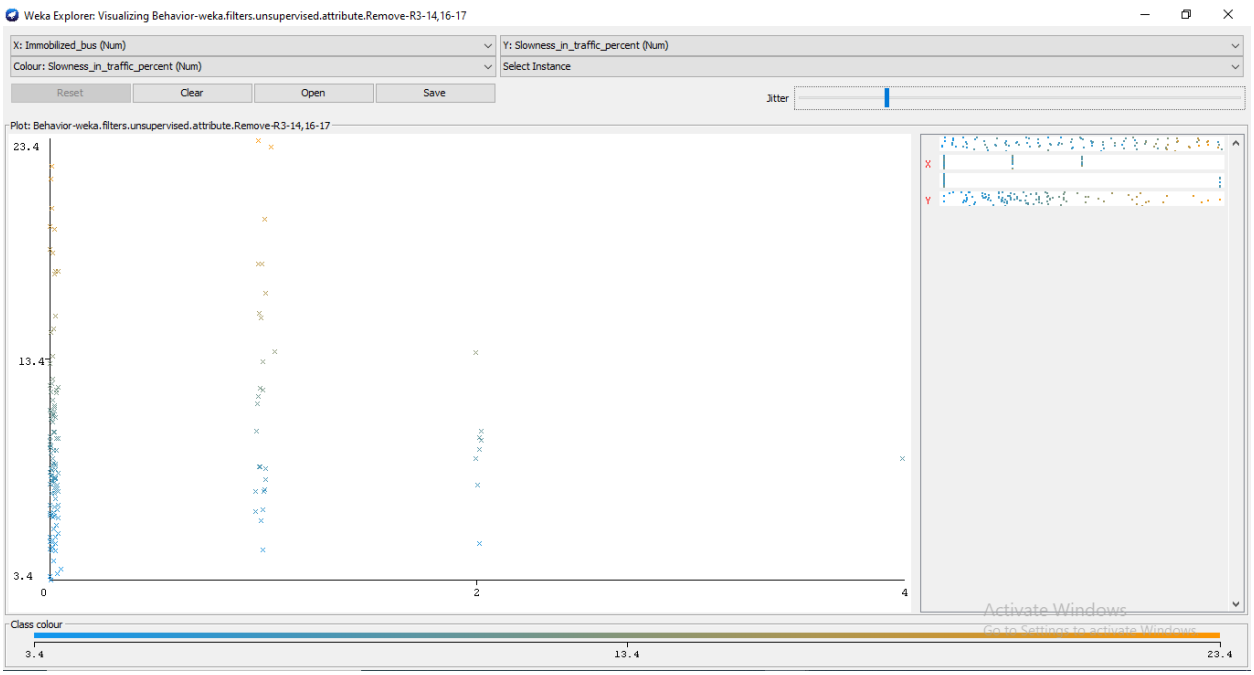


Figure 5.2 Immobilized Bus vs Slowness in Traffic

Figure 5.2, is a plot of behavior of traffic in the city based on x axis as slowness in traffic and y axis as Immobilized bus.

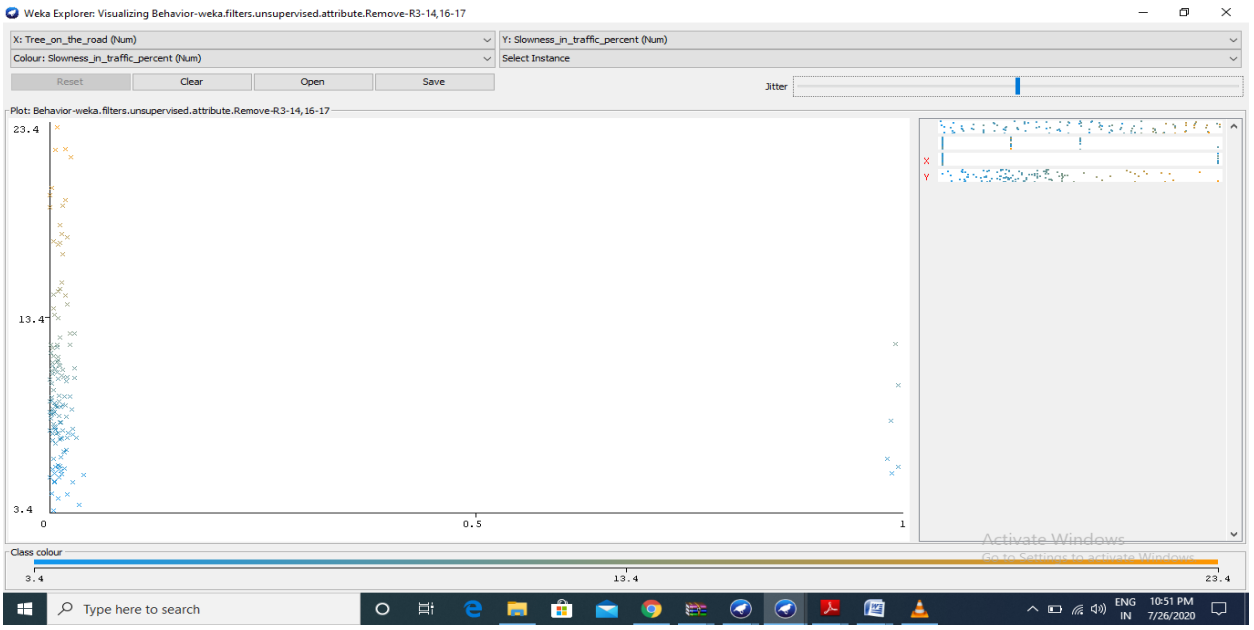


Figure 5.3 Trees On Road vs Slowness In Traffic

Figure 5.3, is a plot of behavior of traffic in the city based on x axis as slowness in traffic and y axis as trees on road.

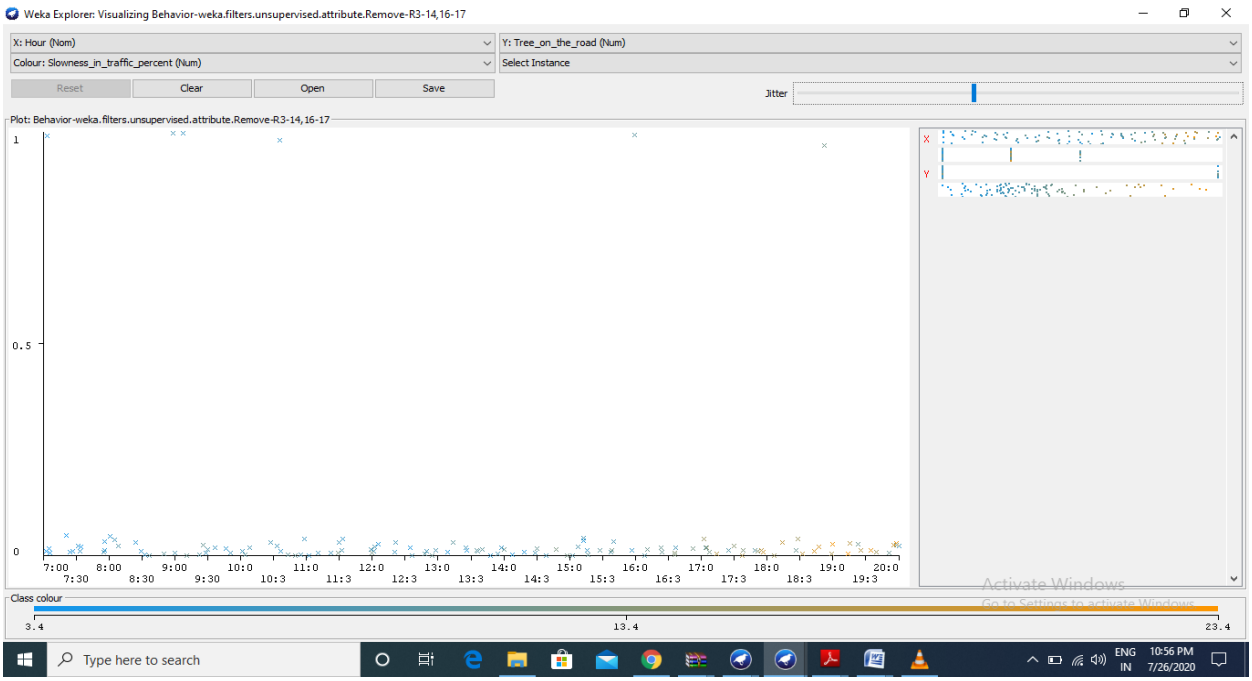


Figure 5.4 Tree On Road vs Hours

Figure 5.4, is a plot of behavior of traffic in the city based on x axis as hours slot and y axis as trees on road.

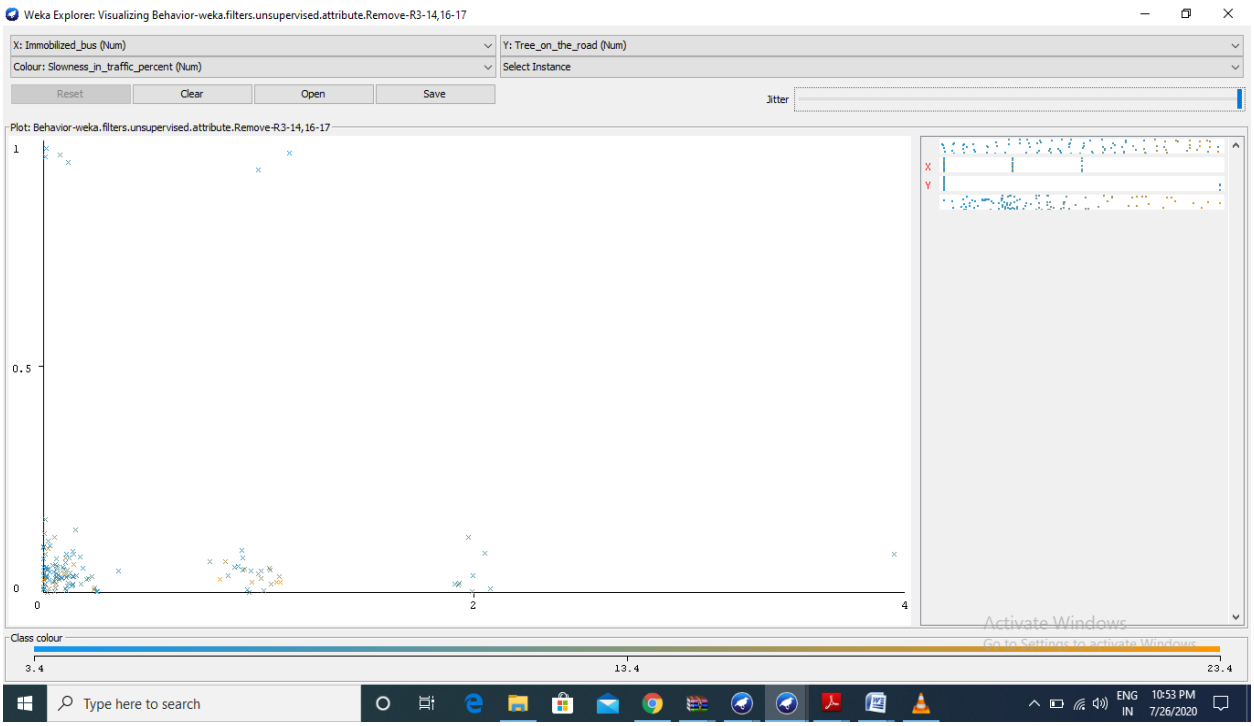


Figure 5.5 Immobilized Bus vs Tree On Road

Figure 5.5, is a plot of behavior of traffic in the city based on x axis as Immobilized bus and on y axis Tree on Road

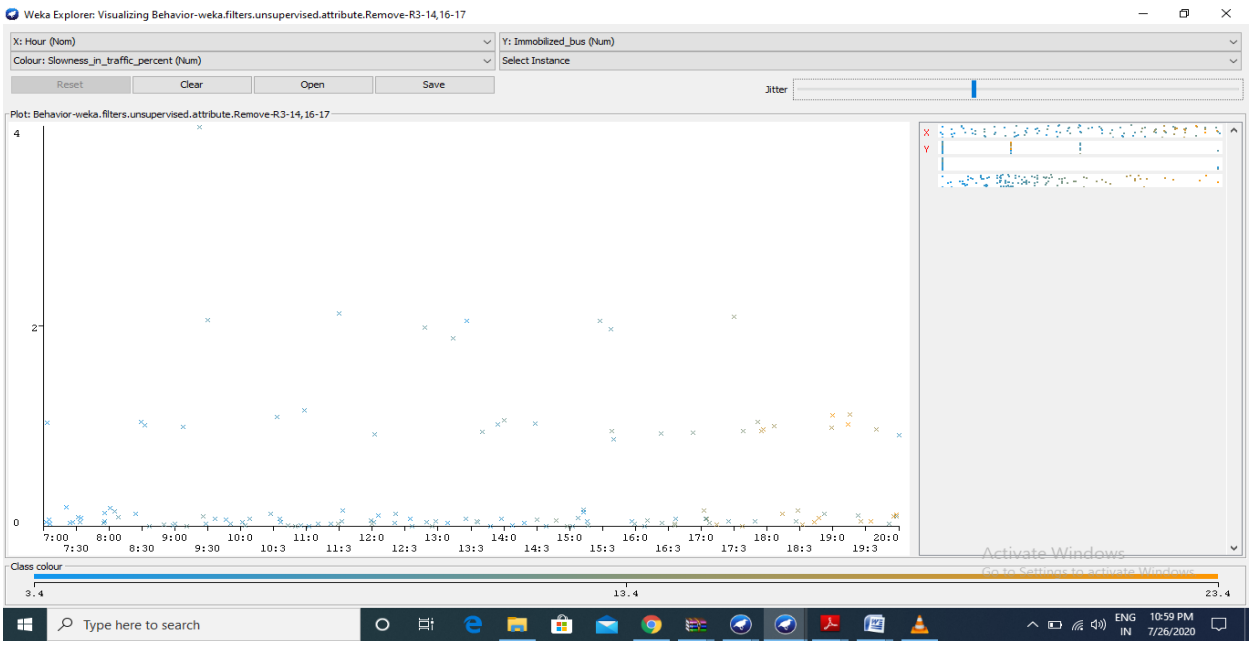


Figure 5.6 Hours vs Immobilized Bus

Figure 5.6, is a plot of behavior of traffic in the city based on x axis as Immobilized bus and on y axis hours slot.

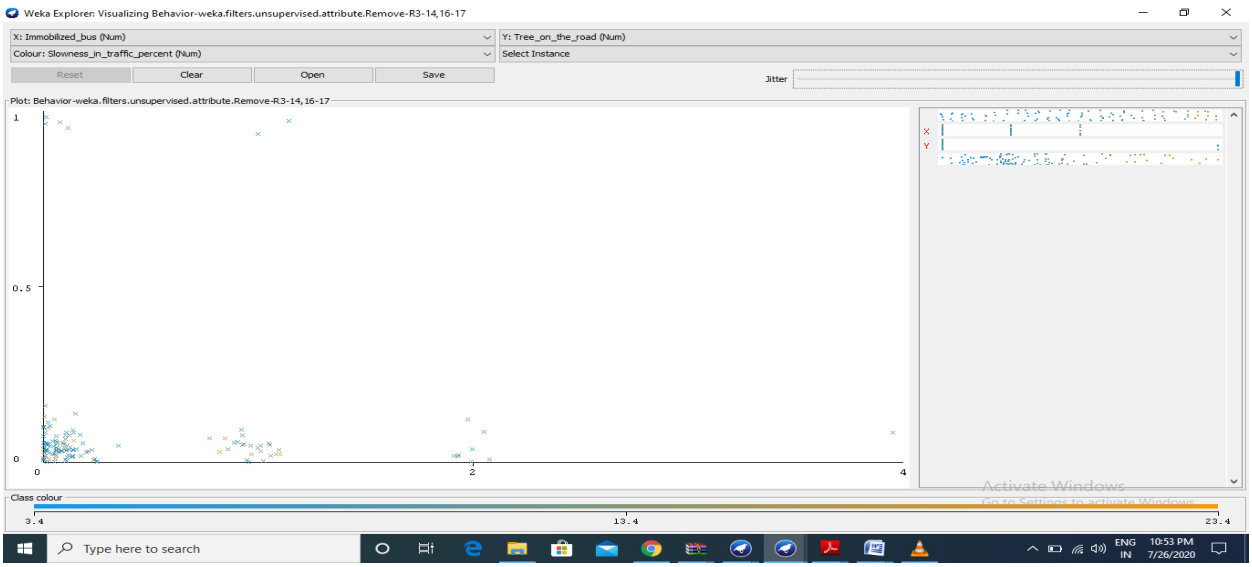


Figure 5.7 Tree On Road vs Immobilized Bus

Figure 5.7, is a plot of behavior of traffic in the city based on x axis as Immobilized bus and on y axis Tree on Road

5.2 CONCLUSION

In a smart city, Big data assumes a significant job in preparing information gathered through IoT gadgets so further examination can be made to perceive the examples and requirements in the city. Sensors introduced everywhere throughout the city produce immense measures of information, yet in the event that it is utilized viably, there are numerous upgrades that should be possible. The prediction of traffic behavior can contribute to decision making prior to routing to support the stages of physical conveyance with more noteworthy adequacy and profitability. With the likelihood to anticipate the vacillations of traffic stream, it is conceivable pick the best windows administration so as to stay away from times when traffic determining point to level of the best.

CHAPTER: 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

The big data issue emerging from the development of a savvy city isn't just a bleeding edge issue for cutting edge logical examination yet in addition a thought process that drives the improvement of a keen city. It brings new chances and difficulties and urges us to facilitate innovation advancement and examination identified with large information, in this way advancing and quickening the improvement of the insight administration industry, better executing the different knowledge applications in a brilliant city, considering more and better shrewd applications to serve the whole society, and making city tasks increasingly logical, proficient, low-carbon and safe. The critical increment in related devices in urban territories has provoked the speedy improvement of data, which has roused the thought of various investigators in different assessment spaces. This research intends to provide a long-term perspective on the work of the great deal of information in a curious city. In this particular case, we examine the powerful innovations are used in a curious city. Distinctive contextual analyses were additionally illustrated. At long last, a few open exploration challenges were disclosed to give the examination bearings to the new analysts in the area. Lastly, we infer that big data can assume a significant job regarding increasing important data and for dynamic purposes.

6.2 FUTURE WORK

In a smart city, Big data assumes a significant job in preparing information gathered through IoT gadgets so further examination can be made to perceive the examples and requirements in the city. Sensors introduced everywhere throughout the city produce immense measures of information, yet in the event that it is utilized viably, there are numerous upgrades that should be possible. The prediction of traffic behavior can contribute to decision making prior to routing to support the stages of physical conveyance with more noteworthy adequacy and

profitability. With the likelihood to anticipate the vacillations of traffic stream, it is conceivable pick the best windows administration so as to stay away from times when traffic determining point to level of the best.

REFERENCES

1. https://medium.com/dataseries/big-data-and-smart-cities-why-we-need-them-now-a194b2498fb1 The Role of Big Data in Smart City. Available from: https://www.researchgate.net/publication/301803005_The_Role_of_Big_Data_in_Smart_City [accessed Jul 21 2020].
2. Abberley, L., Gould, N., Crockett, K., & Cheng, J. (2017). Modelling road congestion using ontologies for big data analytics in smart cities. In 2017 international smart cities conference (ISC2) (pp. 1–6).
3. Aguilera, U., Peña, O., Belmonte, O., & López-de Ipiña, D. (2017). Citizen-centric data services for smarter cities. Future Generation Computer Systems, 76, 234–247.
4. Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. Journal of Internet Services and Applications, 6(1), 25.
5. Anisetti, M., Ardagna, C., Bellandi, V., Cremonini, M., Frati, F., & Damiani, E. (2018). Privacy-aware big data analytics as a service for public health policies in smart cities. Sustainable Cities and Society, 39, 68–77.
6. Antonic, A., Marjanovic, M., Pripužic, K., & Žarko, I. P. (2016). A mobile crowd sensing ecosystem enabled by cupus: Cloud-based publish/- subscribe middleware for the internet of things. Future Generation Computer Systems, 56, 607–622.
7. Babar, M., & Arif, F. (2017). Smart urban planning using big data analytics to contend with the interoperability in internet of things. Future Generation Computer Systems, 77, 65–76.
8. Bokhari, S., & Saiz, A. (2018). Urban big data: City management and real estate markets (Tech. Rep.). GovLab. Barns, S. (2018). Smart cities and urban data platforms: Designing interfaces for smart governance. City, Culture and Society, 12, 5–12.

9. Bassoo, V., Ramnarain-Seetohul, V., Hurbungs, V., Fowdur, T. P., & Beeharry, Y. (2018). Big data analytics for smart cities. In N. Dey, A. E. Hassanien, C. Bhatt, A. S. Ashour, & S. C. Satapathy (Eds.), Internet of things and big data analytics toward next-generation intelligence (pp. 359–379). Cham, Switzerland: Springer International Publishing.
10. Bibri, S. E. (2018). The iot for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. Sustainable Cities and Society, 38, 230–253.
11. Bibri, S. E., & Krogstie, J. (2017). Smart sustainable cities of the future: An extensive interdisciplinary literature review. Sustainable Cities and Society, 31, 183–212.
12. Blumenstock, J. E. (2016). Fighting poverty with data. Science, 353(6301), 753–754. https://doi.org/10.1126/science.aah5217 Bolívar, M. P. R. (2015). Smart cities: Big cities, complex governance? In Transforming city governments for successful smart cities (pp. 1–7). Springer.
13. Chatterjee, S., & Kar, A. K. (2015). Smart cities in developing economies: A literature review and policy insights. In 2015 international conference on advances in computing, communications and informatics (ICACCI) (pp. 2335-2340).
14. Chauhan, S., Agarwal, N., & Kar, A. K. (2016). Addressing big data challenges in smart cities: A systematic literature review. info, 18(4), 73–90. https://doi.org/10.1108/info-03-2016-0012
15. Digiesi, S., Facchini, F., Mossa, G., Mummolo, G., & Verriello, R. (2015). A cyber-based dss for a low carbon integrated waste management system in a smart city. IFAC-PapersOnLine, 48(3), 2356–2361. https://doi.org/10.1016/j.ifacol.2015.06.440
16. Docherty, I., Marsden, G., & Anable, J. (2018). The governance of smart mobility. Transportation Research Part A: Policy and Practice, 115, 114–125.

17. Druedah, J., & Munk-Nielsen, A. (2018). Higher-order income dynamics with linked regression trees. In 2nd conference on dynamic structural models: Methodology and applications of structural dynamic models and machine learning. Department of Economics, University of Copenhagen.
18. TODAY Logistics & Supply Chain. São Paulo: Cecilia Borges, Ano III, n. 38, 2009.
19. Docherty, I., Marsden, G., & Anable, J. (2018). The governance of smart mobility. Transportation Research Part A: Policy and Practice, 115, 114–125.
20. Abberley, L., Gould, N., Crockett, K., & Cheng, J. (2017). Modelling road congestion using ontologies for big data analytics in smart cities. In 2017 international smart cities conference (ISC2) (pp. 1–6).
21. CUNHA, C. B. Aspectos Práticos da Aplicação de Modelos de Roteirização de Veículos a Problemas Reais. Departamento de Engenharia de transportes Escola Politécnica da Universidade de São Paulo, 1997.
22. Zhang, Q., Zhang, X., Zhang, Q., Shi, W., & Zhong, H. (2016). Firework: Big data sharing and processing in collaborative edge environment. In Hot topics in web systems and technologies (hotweb), 2016 fourth IEEE workshop on (pp. 20–25).
23. BALLOU, R. H. Logística Empresarial: transportes, administração de materiais e distribuição física. Tradução de Hugo T. Y. Yoshizaki. São Paulo: Atlas, 1993.
24. CET - Companhia de Engenharia de Tráfego. Disponível em: http://www.cetsp.com.br Acesso em: 16 Dez. 2009.
25. PENA, F. Biografias em fractais: múltiplas identidades em redes flexíveis e inesgotáveis. Revista Fronteiras – estudosmidiáticos, Vol. VI n. 1, p. 82 - jan/jun. 2004.
26. FERREIRA, R.P.; AFFONSO, C.; SASSI, R. J. Roteirização dinâmica de veículos combinada à Previsão do comportamento do tráfego urbano utilizando uma rede neuro fuzzy. In: IV Workshop de Tecnologia Adaptativa (WTA 2010), 2010.

27. FERREIRA, R. P.; AFFONSO, C.; SASSI, R. J. Application of a neuro fuzzy network to forecast the behavior of the urban traffic of the city of São Paulo. In: IEEE Proceedings of Conference Information Systems and Technologies (CISTI), 2010 5th Iberian Conference on, Santiago de Compostela, España, Junio 16-19, pp. 82-85, 2010.
28. NICOLETTI, M. C., CAMARGO, H.A. Fundamentos da Teoria de Conjuntos Fuzzy, Edusfscar, 2004.
29. https://archive.ics.uci.edu/ml/datasets/Metro+Interstate+Traffic+Volume
30. Zaman, H. B., Ahmad, A., Hamid, N. A., Sin, A. K., Hussain, A., Hannan, M. A., & Saad, H. M. (2017). Integrated smart neighborhood framework and application to sustain an innovative digital economy in the 4ir and big data era. In 2017 3rd international conference on science in information technology (ICSITECH) (pp. 7–12).
31. Zaouali, K., Rekik, R., & Bouallegue, R. (2018). Deep learning forecasting based on auto-lstm model for home solar power systems. In 2018 IEEE 20th international conference on high performance computing and communications; IEEE 16th international conference on smart city; IEEE 4th international conference on data science and systems (hpcc/smartcity/dss) (pp. 235–242).
32. Soomro K, Bhutta MNM, Khan Z, Tahir MA. Smart city big data analytics: An advanced review. WIREs Data Mining Knowl Discov. 2019;e1319. https://doi.org/10.1002/widm.1319
33. Pradeep Pai T, Shashikala K L, “Smart City Services - Challenges and Approach" published in IEEE 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con)
34. Lasse Berntzen, Marius Rohde Johannessen, Rania El-Gazzar “Smart Cities, Big Data and Smart Decision- making Understanding "Big Data" in Smart City Applications"

published in ResearchGate ICDS 2018 : The Twelfth International Conference on Digital Society and eGovernments

35. Hamid Fekri Azgomi, Mo Jamshidi, “A Brief survey on Smart Community and Smart Transportation" published in 2018 IEEE 30th International Conference on Tools with Artificial Intelligence
36. Sanjeev Kumar, Anand Prakash, “Role of Big Data and Analytics in Smart Cities" published in International Journal of Science and Research (IJSR)
37. Ibrahim Abaker Targio Hashem, Victor Chang, Nor Badrul Anuar, Kayode Adewole, Ibrar Yaqoob, Abdullah Gani, Ejaz Ahmed, Haruna Chiroma, "The role of big data in smart city" published in International Journal of Information Management 36(5) October 2016
38. Cheng, B., Longo, S., Cirillo, F., Bauer, M., & Kovacs, E, "Building a Big Data Platform for Smart Cities: Experience and Lessons from Santander" published in 2015 IEEE International Congress on Big Data.
39. Kitchin, R. (2014). Making sense of smart cities: addressing present shortcomings. Cambridge Journal of Regions, Economy and Society, 8(1), 131–136.
40. Sumedha Chauhan Neetima Agarwal Arpan Kumar Kar , (2016),"Addressing Big Data Challenges in Smart Cities: A Systematic Literature Review", info, Vol. 18 Iss 4 pp.
41. Gang Zeng, "Application of Big Data in Intelligent Traffic System" published in 2015 IOSR Journal of Computer Science (volume 17, issue 1, ver.vi) PP 01-04.
42. YANG Zhao-xia, ZHU Ming-hua, "A Dynamic Prediction Model of Real-Time Link Travel Time Based on Traffic Big Data" published in 2019 International Conference on Intelligent Transportation, Big Data and Smart City.

43. V.M. Vidya, N. Deepa "Big Data Analytics in Intelligent Transportation Systems using Hadoop" published in 2019 International Journal of Recent Technology and Engineering IJRTE.
44. Feras M. Awaysheh, Jose C. Cabaleiro, Tomas F. Pena, Mamoun Alazab "Big Data Security Frameworks Meet the Intelligent Transportation Systems Trust Challenges" published in 2019 IEEE International Conference on Trust, Security and Privacy in Computing and Communication/13th IEEE International Conference on Big Data Science and Engineering.
45. Nawal Alsaffar, Dr. Hayat Ali and Dr. Wael Elmedany, "Smart Transportation System:A Review of Security and Privacy Issues" published in 2018 3ICT International Conference on Innovation and Intelligence for Informatics, Computing and Technologies.
46. Liu Yang, "Big Data Technology and Its Analysis of Application in Urban Intelligent Transportation System" published in 2018 International Conference on Intelligent Transportation, Big Data and Smart City.
47. P. Sui, X. Li and Y. Bai, "A Study of Enhancing Privacy for Intelligent Transportation Systems: k-Correlation Privacy Model Against Moving Preference Attacks for Location Trajectory Data," IEEE, vol. 5, pp. 24555-24567, 2017
48. F. Kargl, A. Friedman and R. Boreli, "Differential Privacy in Intelligent Transportation Systems," in WiSec '13 Proceedings of the sixth ACM Conference on Security and Privacy in Wireless and Mobile Networks., New York, 2013.
49. J. Lin, J. Niu, H. Li and M. Atiquzzaman, "ASecureandEfficientLocation-basedServiceSchemeforSmart Transportation," FutureGenerationComputerSystems, pp. 1-11, 2017

50. V. Sucasas, G. Mantas, F. Saghezchi, A. Radwan and J. Rodriguez, "An autonomous privacy-preserving authentication scheme for intelligent transportation systems," computers & security, vol. 60, pp. 193-205, 2016.

PLAGIARISM CHECK REPORT

Document Information

Analyzed document	Thesis Report.docx (D77674184)
Submitted	8/11/2020 2:26:00 PM
Submitted by	Naseem
Submitter email	mnaseem@iul.ac.in
Similarity	14%
Analysis address	mnaseem.iul@analysis.urkund.com

Sources included in the report

W	URL: https://www.researchgate.net/publication/315883071_The_role_of_big_data_cloud_comp ... Fetched: 5/30/2020 2:42:31 PM	1
W	URL: https://medium.com/dataseries/big-data-and-smart-cities-why-we-need-them-now-a194b ... Fetched: 8/11/2020 2:27:00 PM	10
W	URL: https://eprints.soton.ac.uk/394830/1/IJIM_Big_Data_in_Smart_City_Final.pdf Fetched: 10/1/2019 10:01:40 AM	12
W	URL: https://www.thinkmind.org/download.php?articleid=intsys_v12_n12_2019_3 Fetched: 8/11/2020 2:27:00 PM	4
W	URL: https://www.ijsr.net/get_abstract.php?paper_id=NOV161007 Fetched: 8/11/2020 2:27:00 PM	3
W	URL: https://doi.org/10.1108/info-03-2016-0012 Fetched: 8/11/2020 2:27:00 PM	4
SA	**Computer Science & Engineering Revati Machhindra Wahul Ph. D Thesis.pdf** Document Computer Science & Engineering Revati Machhindra Wahul Ph. D Thesis.pdf (D59971897)	2
W	URL: https://imacimi.wordpress.com/2019/09/03/big-data-and-smart-cities-how-prepared-ar ... Fetched: 8/11/2020 2:27:00 PM	1
W	URL: https://www.researchgate.net/publication/333871841_Smart_city_big_data_analytics_A ... Fetched: 11/16/2019 8:05:27 AM	4
W	URL: https://d-nb.info/1149998245/34 Fetched: 8/11/2020 2:27:00 PM	1

Entire Document

ABSTRACT

Welcome to the era of either big data, as we all know this is the era of big data and every model in this era is directly or indirectly by utilizing the concepts of big data. Most important aspect of big data is to handle volume, velocity and Variety. While implementing the concept of smart city these features of big data will play an important role to efficiently manage various services. Transportation can be handily overseen with the aid of Big Data. A jam or a congestion on road is a very weary thing for commuters. The analysis of vehicular movement related data work can be done to address this issue effectively. Big Data will also serve this purpose. In our paper we are going to prepare, a data analytics-based model to tackle all these problems. For this we will utilize the techniques of big data because traffic related data segment in any urban city has voluminous in nature and size of data will decrease as traffic density will increase. The aim of this paper is to estimate the conduct of the urban traffic of any urban city. The strategy of the paper comprises in the catch of important occasions that influence the progression of traffic of the city and the utilization of information investigation to prepare with these prominent events to foresee the conduct of traffic. This paper presents the outcomes in anticipating the conduct of the traffic of the city for seven days.

CHAPTER: 1

INTRODUCTION

1.1 INTRODUCTION OF BIG DATA

Big Data is additionally data but with an enormous size. Big Data may be a term won't to describe a set of knowledge that's huge in volume and yet growing exponentially with time. In short, such data is so large and sophisticated that none of the normal data management tools are ready to store it or process it efficiently.

Simply, big data is very big, more complex data, especially from new data sources. These data sets are so voluminous that traditional processing software just can't manage them. But these massive volumes of knowledge are often wont to address business problems you wouldn't are ready to tackle before.

In big data, data contains greater variety outcomes in rapidly increasing volumes and with higher velocity. This is known as the three Vs. 1.1.1 The Three Vs of Big Data Volume – The name Big Data itself is said to a size which is gigantic. Size of data plays a very crucial role in determining value out of data. Also, whether a specific data can actually be considered as an enormous Data or not, depends upon the quantity of knowledge. Hence, 'Volume' is one characteristic which must be considered while handling Big Data.

Velocity – The term 'velocity' refers to the speed of generation of knowledge. How fast the info is generated and processed to satisfy the stress, determines real potential within the data. Velocity of data in big data means the speed at which data flows in from of sources like businesses, applications, network areas, sensors,

75%	MATCHING BLOCK 1/42	W	https://www.researchgate.net/publication/ ...

mobile devices, etc. The flow of data is massive and continuous.
Variability – It refers to the

inconsistency which can be shown by the data at times, so hampering the process of being able to handle and manage the data effectively and easily. There are two more Vs have emerged over the past few years: Value and Veracity. Data has intrinsic value. But it's of no use until that value is known and knowledgeable. Equally important: How truthful is your data —and how much can you rely on it? Today, big data has become capital. Think of some of the world's biggest tech companies. A large and important part of the value they offer comes from their business data, which they're constantly analyzing to produce more efficiency and develop new products. With an increased volume of massive data now cheaper and more accessible, you'll make more accurate and precise business decisions. Finding value in big data isn't only about analysing it (which is a whole other benefit). It's an entire discovery process that requires insightful analysts, business users, and executives who ask the right questions, recognize patterns, make informed assumptions, and predict behaviour. 1.1.4 The History of Big Data Although the concept of massive data itself is comparatively new, the origins of huge data sets return to the 1960s and '70s when the planet of knowledge was just getting started with the first data centers and the

development of the relational database. Around 2005, people began to understand just what proportion data users generated through Facebook, YouTube, and other online services. Hadoop (an open-source framework created specifically to store and analyses big data sets) was developed that same year. NoSQL also began to realize popularity during this point. The development of open-source frameworks, like Hadoop (and more recently, Spark) was essential for the expansion of massive data because they create big data easier to figure with and cheaper to store. In the years since then, the quantity of massive data has skyrocketed. Users are still generating huge amounts of data—but it's not just humans who do it. With the advent of the Internet of Things (IoT), more objects and devices are connected to the internet, gathering data on customer usage patterns and product performance. The emergence of machine learning has produced still more data. While big data has come far, its usefulness is only just beginning. Even further with cloud computing expanded possibilities of big data. The cloud offers truly elastic scalability, where developers can simply spin up unplanned clusters to check a subset of knowledge. 1.1.5 Parts of Big Data Big Data' might be found in three forms: i. Structured ii. Unstructured iii. Semi-structured Structured Any data which will be stored, accessed and processed within the sort of fixed format is termed as a 'structured' data. Over the amount of your time, talent in computing has achieved greater success in developing techniques for working with such quite data (where the format is documented in advance) and also deriving value out of it. However, nowadays, we are foreseeing issues when a size of such data grows to an enormous extent, typical sizes are being within the rage of multiple zettabytes. Unstructured Any data with unknown form or the structure is assessed as unstructured data. additionally, to the dimensions being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it. A typical example of unstructured data may be a heterogeneous data source containing a mixture of straightforward text files, images, videos etc. Now day organizations have wealth of knowledge available with them but unfortunately, they do not skills to derive value out of it since this data is in its raw form or unstructured format. Semi-structured Semi-structured data can contain both the sorts of data. we will see semi-structured data as a structured in form but it's actually not defined with e.g. a table definition in relational DBMS. Example of semi-structured data may be a data represented in an XML file. 1.1.6 Benefits of massive Data and Data Analytics: • Big data makes it possible for you to realize more complete answers because you've got more information. • More complete answers mean more confidence within the data—which means a totally different approach to tackling problems. 1.1.7 Big Data Challenges While big data holds tons of promise, it's not without its challenges. First, big data is... big. Although new technologies are developed for data storage, data volumes are doubling in size about every two years. Organizations still struggle to stay pace with their data and find ways to effectively store it. But it's not enough to only store the info. Data must be wont to be valuable which depends on curation. Clean data, or data that's relevant to the client and arranged during a way that permits meaningful analysis, requires tons of labor. Data scientists spend a lot of their time curating and preparing data before it can actually be used. Finally, big data technology is changing at rapidly. a couple of years ago, Apache Hadoop was the favored technology won't to handle big data. Then Apache Spark was introduced in 2014. Today, a mixture of the 2 frameworks appears to be the simplest approach. maintaining with big data technology is an ongoing challenge. 1.1.8 How Big Data Works Big data gives you new insights that open up new opportunities and business models. Getting started involves three key actions: i. Integrate Big data brings data together from many sources and applications. Traditional data integration mechanisms, like ETL (extract, transform, and load) generally aren't up to the task. It requires new strategies and technologies to research big data sets at terabyte, or maybe petabyte, scale. During integration, you would like to usher in the info , process it, and confirm it's formatted and available during a form that your business analysts can start with. ii. Manage Big data requires storage. Your storage solution is often within the cloud, on premises, or both. you'll store your data in any form you would like and convey your required processing requirements and necessary process engines to those data sets on an on-demand basis. many of us choose their storage solution consistent with where their data is currently residing. The cloud is gradually gaining popularity because it supports your current compute requirements and enables you to spin up resources as required. iii. Analyze Your investment in big data pays off once you analyze and act on your data. Get new clarity with a visible analysis of your varied data sets. Explore the info further to form new discoveries. Share your findings with others. Build data models with machine learning and AI. Put your data to figure.

1.2 INTRODUCTION OF SMART CITY

inhabitants. There are a wide range of motivations to concentrate on making our urban areas savvy, over half total populace is living in urban communities and there are 60% urban communities around the world that are yet to be assembled. Urban areas are at the vanguard of worldwide development and by 2050, over 70% of our total populace will live in the urban areas. While looking at making our traditional urban communities keen, we ought to likewise consider that shrewd urban communities ought to be resident driven, collective, responsive, practical, responsible and straightforward so as to help individuals in a most useful way.

Smart Technology Healthcare Infrastruture Education Security Buildings Energy Governance and Citizen Transport

Fig. 1. Components of The Smart City

things to come. We can see it through the arrangement of network assets (network

Since information is the new gold, so by utilizing sensors or by deriving information from various devices like mobile phones etc., large database of information can be gathered and Big Data investigation can be utilized on this data to manage the smart structures of a smart city [12]. The smart city enormous information system must ensure an effective stockpiling of the shifted information structures (organized, to un-organized and semi-organized). They should have the ability to process both continuous just as recorded information (various necessities of ongoing versus clump handling). In addition, they should give adaptability regarding information stockpiling and handling (in case of an unexpected increment in load). At long last, they ought to likewise have the option to share the handled outcomes over an assortment of uses/administrations in a gradual and versatile way. 1.2.1 Smart City Applications of Big Data The utilization of big data (enormous information) propels for the keen(smart) city enables compelling data amassing and taking care of to make information that can update various brilliant city organizations. Likewise, huge information enables leaders to get ready for any extension in brilliant (

viable information investigation. 1.2.1.1