JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

VEHICLE DATA ANALYZATION USING POWER BI

[1]Prof. Samita Patil, [2]Akanksha Tambitkar, [3]Rutuja Phadatare, [4]Vaishnavi Pol

Professor (Guide), Student, Student, Student

[1,2,3]Department of Computer Engineering, University of Mumbai, Shivajirao S. Jondhale College of Engineering Maharashtra

Abstract: Autonomous vehicles have been developed to enhance transportation safety by sensing their environment and making decisions without external assistance, ultimately producing an optimal route to reach a destination. They are the envisioned future of smart, efficient, and crash-avoiding urban vehicles. Automakers have begun working in this area to realize the potential and address the current challenges to achieve the expected outcome. However, analyzing the energy consumption, vehicle-specific power, and power consumption poses some difficulties. To overcome these issues, we are creating a big data application called "Vehicles Data Analyzation." This application collects vehicle data several times per second, enabling fine-grained and near-real-time analysis of vehicle status and operating behavior. Our focus is on typical streaming applications, and we present implementations using Power BI. We compare the different architectures of autonomous vehicular systems, including speed, vehicle-specific power, actual power consumption, and energy consumption. We also compare the different databases such as Hadoop and MongoDB. Our ultimate goal is to show the analysis of the fine features integrated into the vehicular system.

Keywords: **- Energy consumption, Vehicle specific power, Power consumption, Performance Evaluation between Hadoop and MongoDB**

I. INTRODUCTION

A fully autonomous vehicle can be defined as a car that can perceive its environment, decide which route to take to its destination and drive it. The number of autonomous vehicles is increasing strongly so we need to do a proper analysis of autonomous vehicle data [1]. In today's world autonomous vehicles generate up to 25 gigabytes of data and this number goes up to 3600 gigabytes every hour. Many useful evaluations can be done from vehicle data to evaluate and improve e.g.: Times (ms), Latitude (deg), Longitu de (deg), Altitude (m), and Vehicle Speed [5].
It is quite difficult to handle and process the enormous volume of data that is generated. To provide data-driven services, it is vital to have computational systems that can deal with this huge volume of data, and analyze and process it to extract meaningful insights [2]. First, we aim to show the potential of fine-grained analysis in terms of how to compute energy consumption, and power consumption, and calculate vehicle-specific power [1]. Secondly, we aim to show the performance evaluation between the databases which include Hadoop and MongoDB. Utilizing the POWER-BI platform, we have examined all of this data. It is a data visualization dashboard that converts data from different sources for Excel workshop, CS file, XML, etc. Power BI helps to excess our data instantly with less manual work. It handles a huge amount of Data and automatically gets connected with it.

II. LITERATURE SURVEY

Big Data Architectures for Vehicle Data Analysis aim to analyze power efficiency and consumption. The analysis is based on Apache Spark and Flink. Currently, there is no proper visualization of vehicle-specific power, and only individual or batch data sets are recorded. When it comes to scalability, Apache Spark batch processing performs better on higher-performance machines compared to the scalability of Flink streaming on the same machines.[1]

Big data for estimating the energy consumption and driving range of electric vehicles. It considers variables such as driving speed and temperature based on the season. However, it does not include a detailed analysis of the architecture of autonomous vehicles. The study focuses solely on the driving range of electric vehicles. The findings of the analysis suggest that the mean ECR of BEVs is 0.183 kW h/km. Furthermore, the findings show that in the winter, compared to the summer, driving range decreases by 25% and ECR increases by 34%. Finally, the study suggests that the electricity tariff for BEVs is cost-efficient compared to conventional vehicles.[2]

Big data analytics architecture for Internet-of-vehicle. The architecture is based on Apache Spark and is used to address issues related to city traffic, traffic safety, and traffic organization. The system analyzes only traffic-related activity, including massive data such as road and traffic flow detection. The paper employs big data analytics and data mining to create a platform for large data analysis of traffic management. [3]

Data analyzation of autonomous electric vehicle on basis of energy consumption of autonomous electric vehicles. The study mainly examines the relationship between a vehicle's power and its speed. The data is divided into different velocity and acceleration ranges. The analysis only considers the energy consumption of the autonomous vehicles and specifically focuses on the battery system. The system that can gather information from electric car batteries, such as battery state of charge, pack voltage, pack power, latitude, and longitude, is presented in the study. The system can also estimate the power of the electric vehicles. [4]

Data Architecture for autonomous vehicles mainly focuses on ensuring the security of the vehicles by analyzing speed data. It generates a massive amount of data, up to 30 GB per day, and performs analysis on speed and temperature. The vehicles use 3G wireless communications. The paper suggests that autonomous vehicles (CVs) can gather a wide range of measurements, including speed and temperature. It is expected that by the end of 2025, there will be approximately 2 billion connected cars on the world's roads, each of which can produce up to 30 gigabytes of data per day. [5]

III. PROBLEM STATEMENT

The problem is when it comes to handling and processing the vast amount of data which is generated by autonomous vehicles. The existing system of autonomous vehicles generates up to only 25 gigabytes of data. It only records the individual data. It does not consider such fine-grained analysis, nor discussed the basic architecture of autonomous vehicles.

Our main objective of the project is to visualize the “AUTONOMOUS VEHICLES DATA”, which would be user-friendly, secure, and easy to visualize and analyze. And also, to show the performance evaluation done by Hadoop and MongoDB databases. The evaluation can be done for example:

Energy Consumption - Calculating the used energy data.

Power Consumption – It is used to calculate the power required by a particular vehicle. Vehicle Specific Power – It estimates the power in kilowatts and time in hours.

Comparison Between – Hadoop and MongoDB

According to the evaluation issues, there should be proper data visualization before buying an autonomous vehicle. Generally, people do not visualize the data properly due to improper data creation. So, data visualization is very important.

IV. METHODOLOGY

Our system focuses on creating a big data application for autonomous vehicles, called Vehicles Data Analyzation. This application receives vehicle data several times per second, allowing for detailed and near real-time analysis of vehicle status and operating behaviors. We concentrate on typical streaming applications and present implementations using Power BI. We compare the different architectures of autonomous vehicular systems which include speed, vehicle specific power, actual power consumption, and energy consumption. Our application uses 3G, 4G wireless technology. Our aim is to present a detailed analysis of the fine features built into vehicular systems and also to show the performance evaluation between different databases such as Hadoop and MongoDB. So, we use the POWER-BI platform for data visualization and database comparison. To achieve this, we use the Power BI platform for data visualization. It is a data visualization dashboard that can convert data from different sources such as Excel workshops, CSV files, and XML files. Power BI helps us to access our data instantly with less manual work.

•Vehicle Edge Computing – It lessens needless data transmission and enables the car to process data. We utilize wireless technologies like 3G and 4G to connect the cloud layer and the smart vehicular layer. Thus, edge computing and vehicle edge computing are related in this way.

•Edge Computing – It refers to the network and device range that is at or close to the user. It is employed to provide a quicker total travel time. A gateway is used to establish a connection between the data Centre and edge computing.

•Gateway - A gateway is a communications network node that joins two networks with dissimilar transmission methods.

•Data Centre – It instantly processes and analyses vast amounts of data, including data from autonomous cars. We used Hadoop and a MongoDB database to hold the data that we extracted from Google and Kaggle.

•Data visualization – It is data and information represented graphically. We use POWER BI as our visualization tool.

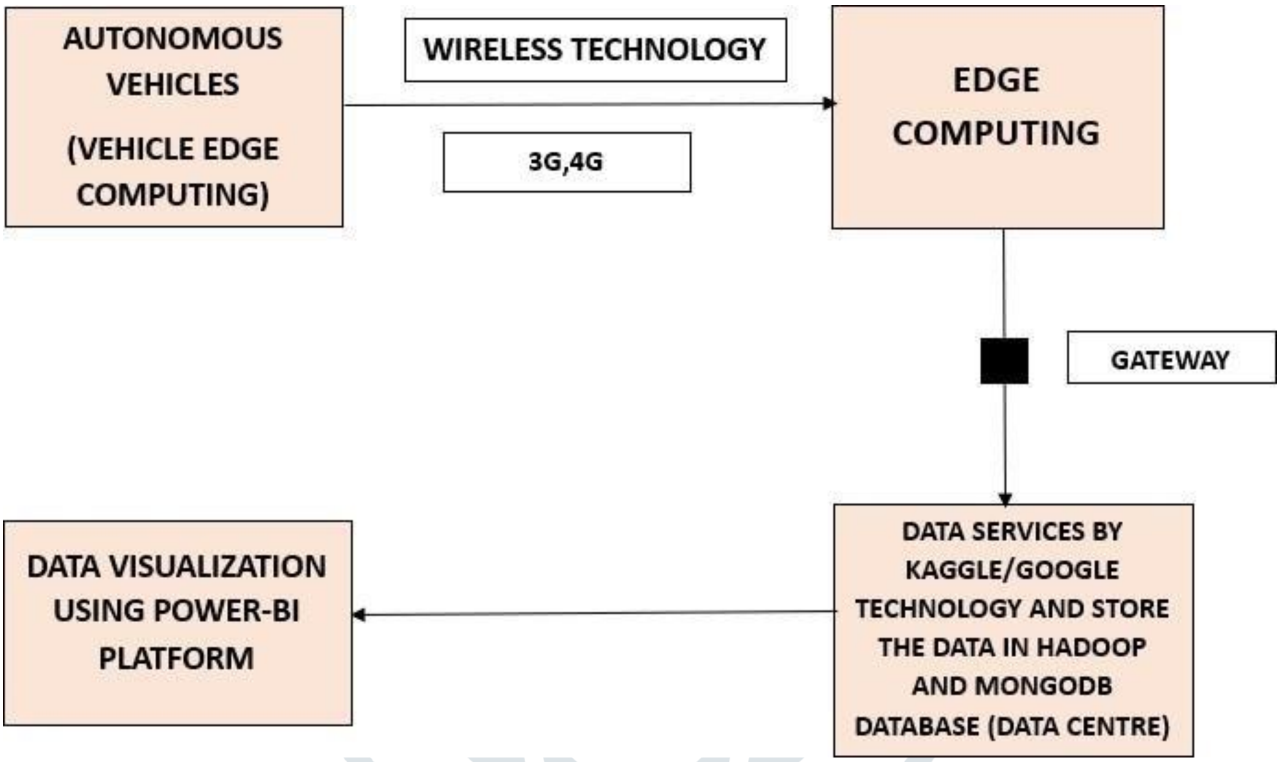

Fig. IV.1. Methodology of System

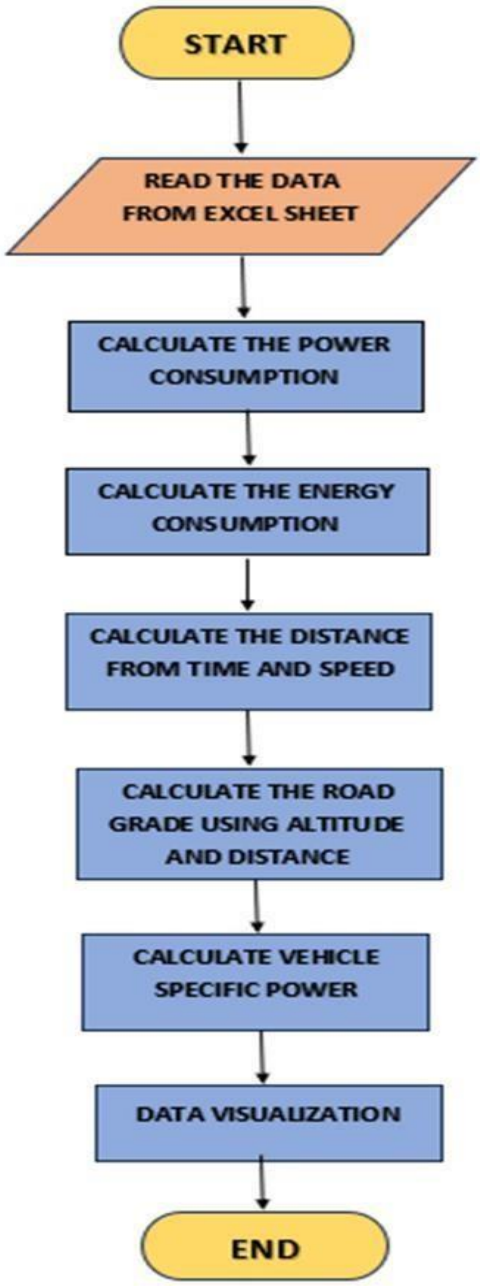

Fig. IV.2. Flowchart of System

V. SYSTEM REQUIREMENTS

Software Used:

- **Microsoft Windows 11 –** It is Windows' most recent operating system version.

- **POWER BI -** This data visualization dashboard transforms data from several sources, such as Excel workshops, CS files, XML files, and so on. With less human labor, Power Bi enables us to immediately extract our data. It is used to extract meaning from an organization's data. With the use of Power BI, it is possible to link unrelated data sets, clean and modify the data into a data model, and produce graphs or charts that show the data visually.

- **Excel Sheet –** It's a Microsoft spreadsheet programmed.

- **CSV FILE –** It is an acronym for COMMA SEPARATED VALUE, a text file format that divides values with commas.

- **Hadoop Database -** Hadoop is an open-source Java-based framework that controls how big data is processed and stored for use in applications. Hadoop divides workloads into smaller tasks that may be completed concurrently by using distributed storage and parallel computing to manage big data and analytics operations.

- **Kaggle** - The largest Data Science community in the world is housed on the web platform Kaggle. With Kaggle, users may discover datasets for AI model construction, share datasets, collaborate with other data scientists and machine learning experts, and participate in competitions to find solutions to data science problems.

- **MongoDB -** A non-relational document database that supports JSON-like storage is called MongoDB. The flexible data model of the MongoDB database makes it possible to store unstructured data. This open-source document-oriented database is made to hold a lot of data and facilitate effective manipulation of that data. Because data in MongoDB is not stored and retrieved as tables, it falls under the category of NoSQL (Not Only SQL) databases.

Hardware Used:

- **Intel Processo**r- The Intel Corporation produces the 2.11GHz Intel® CPU, a particular type of microprocessor. It is the part of a computer's brain that processes commands and data.

- **RAM** – 8GB - Random-access memory is referred to by its common acronym, RAM, in computer terminology. It is sometimes referred to as memory or PC memory. RAM is essentially the short-term memory of your laptop or computer. Your computer processor needs this location of the data in order to run apps and open files.

- **Hard Disk Drive** – 476.96 - The hard drive, also referred to as the hard disc drive, is one type of data storage device that may be found in desktop and laptop computers (HDD). Data that has been stored on an HDD can be kept even if the device is not getting power because it is a "non-volatile" storage disc.

VI. RESULTS

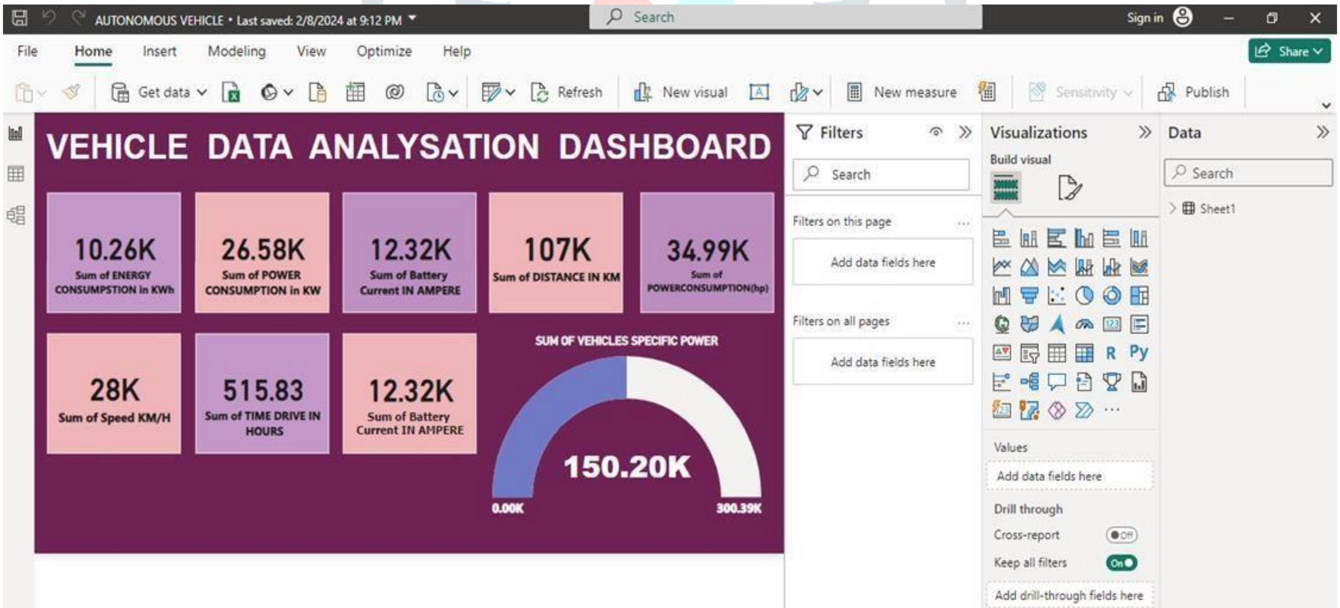

Fig. VI.1. The Key Point Indicators

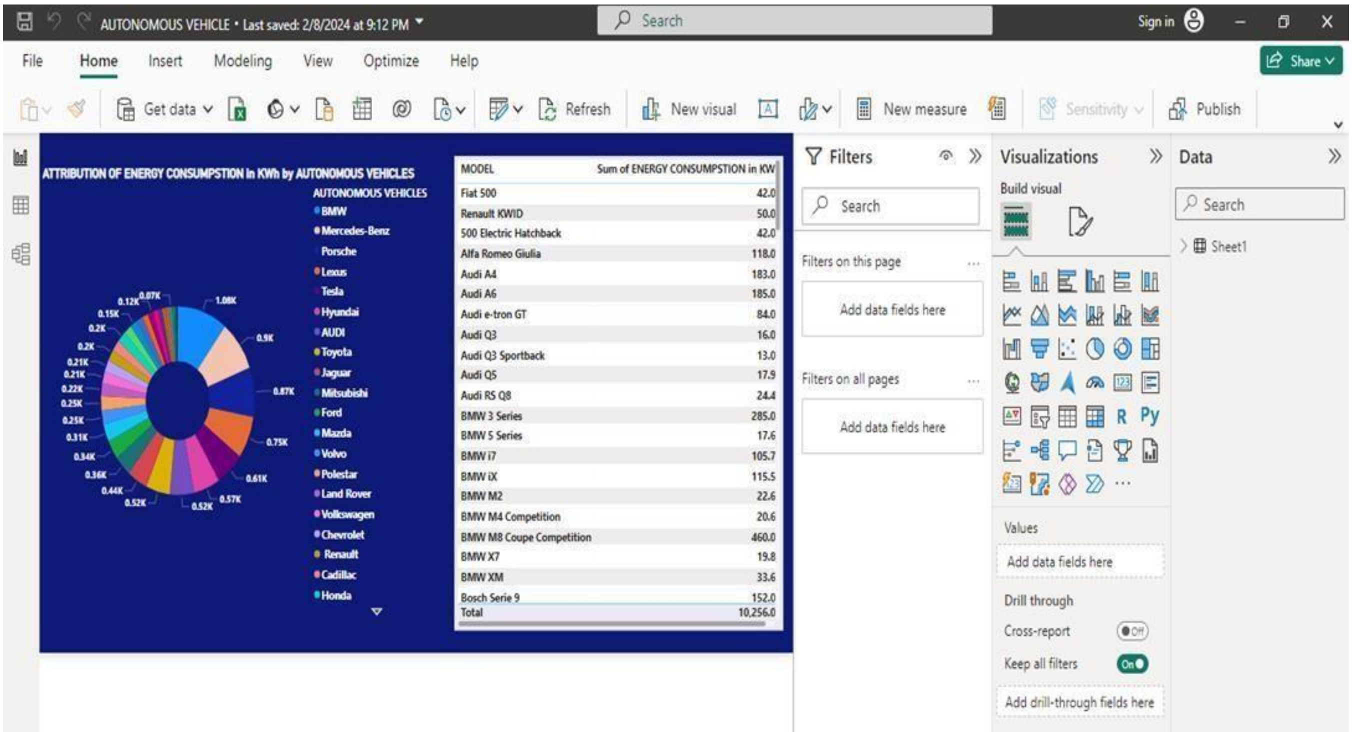

Fig. VI.2. Energy Consumption

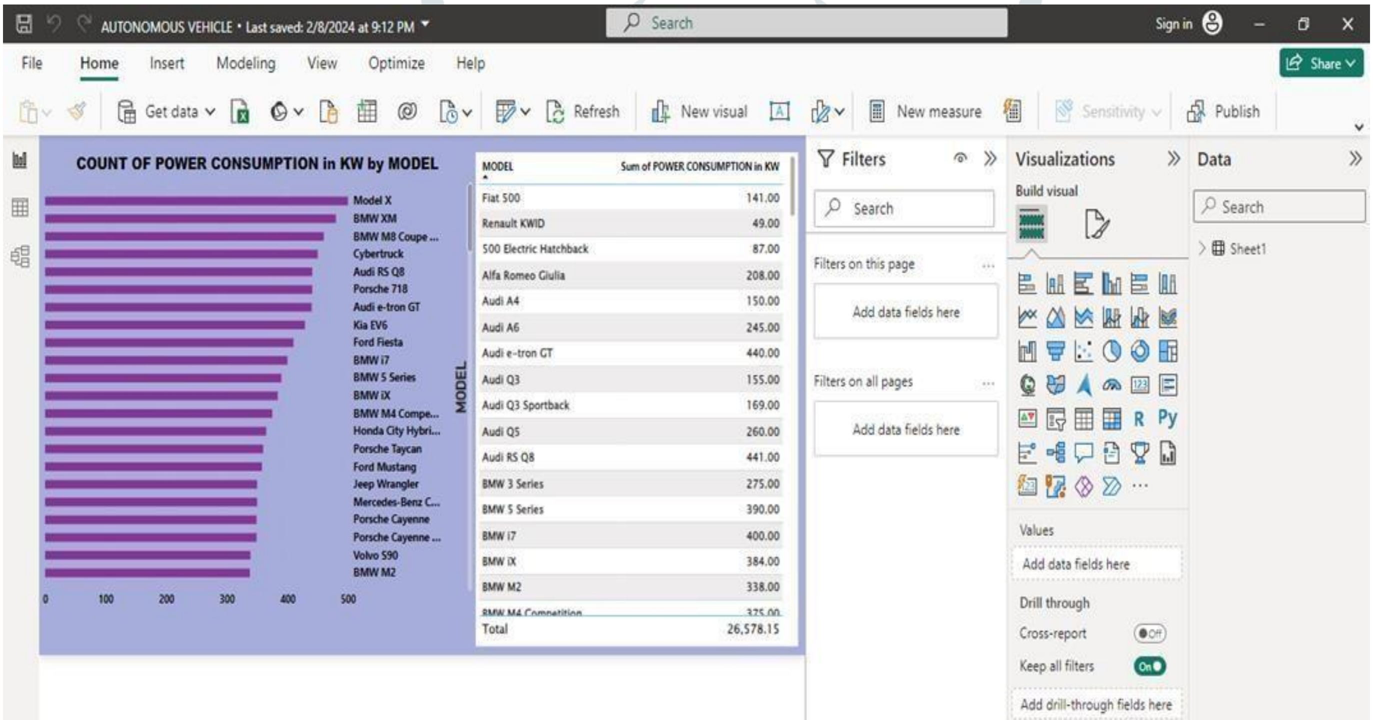

Fig. VI.3. Power Consumption

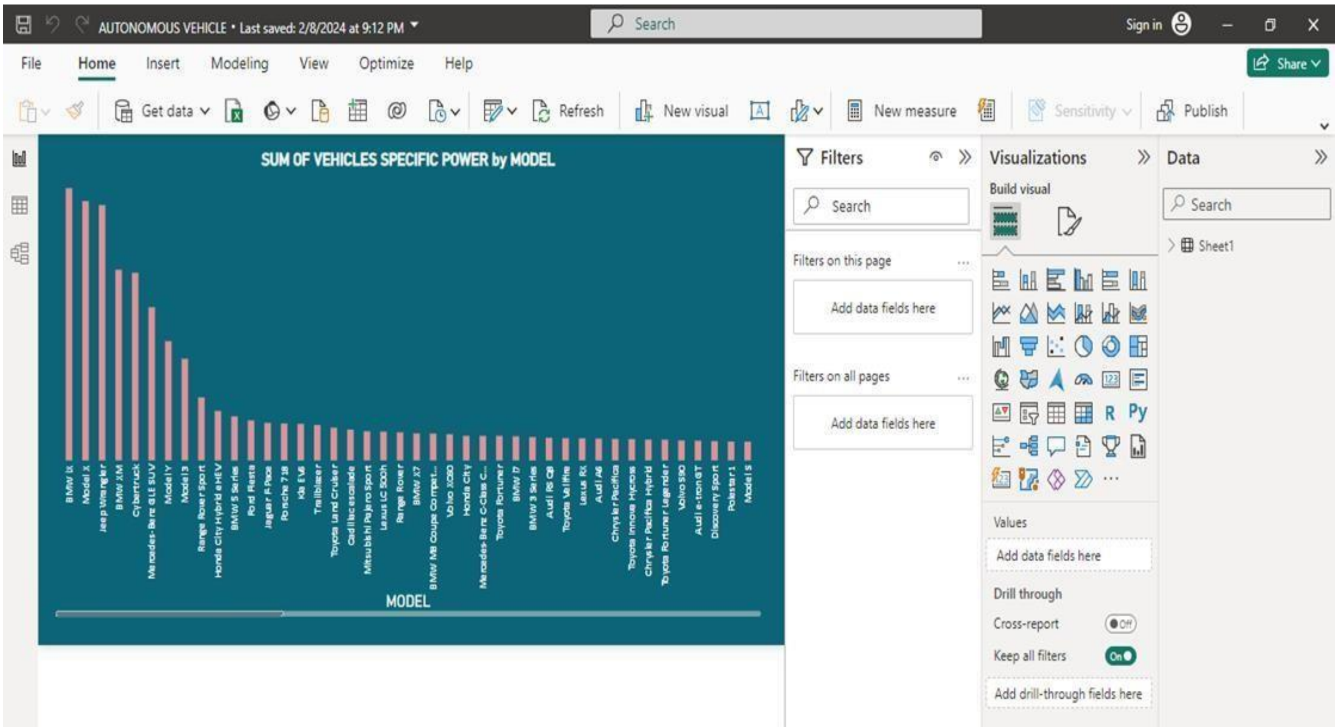

Fig. VI.4. Vehicle Specific Power

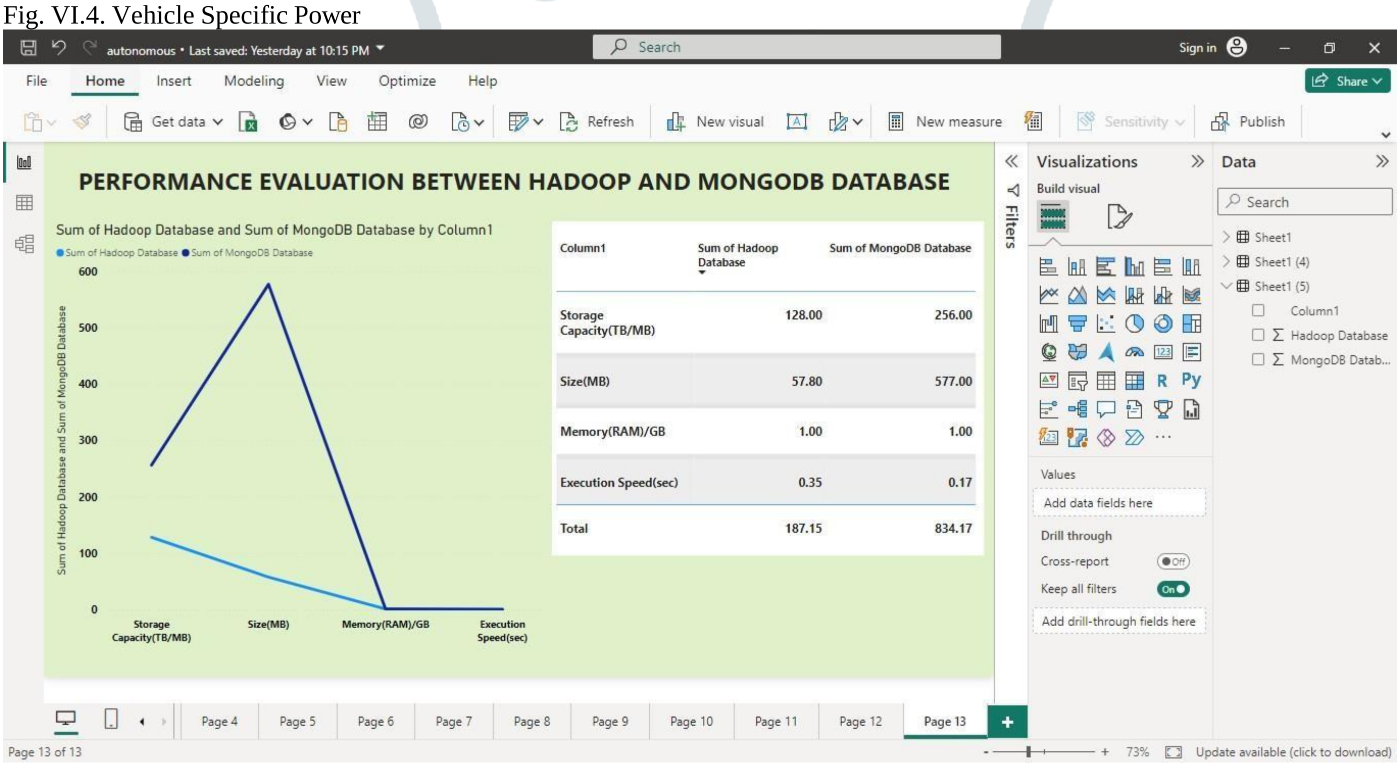

Fig. VI.5.Performance Evaluation Between Hadoop and MongoDB

VII. CONCLUSION

Managing and understanding a large amount of data is challenging, which is why data visualization is crucial for autonomous vehicles. So, our aim to show the analysis of fine-features inbuilt in vehicular system and also to show the performance evaluation between different database such as Hadoop and MongoDB. According to the evaluation issues there should be a proper data visualization before buying an autonomous vehicle.

VIII. ACKNOWLEDEGEMENT

We thank our project guide Prof. Samita Patil along with our Head of Department Dr. Uttara Gogate and our Principal Dr. Pramod R. Rodge for providing us an opportunity of the project work in SSJCOE Dombivli and providing us all support and guidance which made us to pursue the project duly. We are extremely thankful to them for providing such a nice support and guidance. We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of Computer Engineering which helped us in pursuing our project work.

REFERENCES

[1] Christian Prehofer DENSO Automotive Germany Eching, Germany Shafqat Mehmood DENSO Automotive Germany and TU Chemnitz Eching 2020. “Big Data Architectures for Vehicle Data Analysis”.

[2] G. M. Fetene, S. Kaplan, S. L. Mabit, A. F. Jensen, and C. G. Prato 2017. “Big data for estimating the energy consumption and driving range of electric vehicles”.

[3] Dan Liu 2018. “Big data analytics architecture for Internet-of-vehicle Based on the Spark”.

[4] Alfredo Cabrera, William A Kitch David Freese 2017. “Data analyzation of autonomous electric vehicle on basis of energy consumption”.

[5] Anthony Nassar, Amir Haroun, Francois Dessables 2020. “Data Architecture for autonomous vehicles”.

[6] T. Furuichi and K. Yamada, “Next generation of embedded system on cloud computing,” Procedia Computer Science, vol. 35, pp. 1605–1614,12 2014.

[7] S. Teerapittayanon, B. McDanel, and H.-T. Kung, “Distributed deep neural networks over the cloud, the edge and end devices,” in 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2017, pp. 328–339.

[8] D. Battulga, D. Miorandi, and C. Tedeschi, “Fogguru: a fog computing platform based on Apache flink,” in 2020 23rd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN). IEEE, 2020, pp. 156–158.

[9] O.-C. Marcu, A. Costan, G. Antoniu, and M. S. Perez-Hern ´ andez, “Spark ´ versus flink: Understanding performance in big data analytics frameworks,” in 2016 IEEE International Conference on Cluster Computing (CLUSTER). IEEE, 2016, pp. 433–442.

[10] H. Lee, J. Oh, K. Kim, and H. Yeon, “A data streaming performance evaluation using resource constrained edge device,” in 2017 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2017, pp. 628–633.

[11] C.-F. Yeh, L.-T. Lin, P.-J. Wu, and C.-C. Huang, “Using on-board diagnostics data to analyze driving behaviors and fuel consumption,” in International Conference on Smart Vehicular Technology, Transportation, Communication and Applications. Springer, 2018, pp. 343–351.

[12] E. Ericsson, “Independent driving pattern factors and their influence on fuel-use and exhaust emission factors,” Transportation Research Part D: Transport and Environment, vol. 6, no. 5, pp. 325–345, 2001.

[13] C. De Cauwer, W. Verbeke, T. Coosemans, S. Faid, and J. Van Mierlo, “A data-driven method for energy consumption prediction and energy efficient routing of electric vehicles in real-world conditions,” Energies, vol. 10, no. 5, p. 608, 2017.

[14] O.-C. Marcu, A. Costan, G. Antoniu, and M. S. Perez-Hern ´ andez, “Spark ´ versus flink: Understanding performance in big data analytics frameworks,” in 2016 IEEE International Conference on Cluster Computing (CLUSTER). IEEE, 2016, pp. 433–442.

[15] H. Lee, J. Oh, K. Kim, and H. Yeon, “A data streaming performance evaluation using resource constrained edge device,” in 2017 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2017, pp. 628 –633.

[16] C.-F. Yeh, L.-T. Lin, P.-J. Wu, and C.-C. Huang, “Using on-board diagnostics data to analysis driving behaviors and fuel consumption,” in International Conference on Smart Vehicular Technology, Transportation, Communication and Applications. Springer, 2018, pp. 343–351.

[17] E. Ericsson, “Independent driving pattern factors and their influence on fuel-use and exhaust emission factors,” Transportation Research Part D: Transport and Environment, vol. 6, no. 5, pp. 325–345, 2001.

[18] C. De Cauwer, W. Verbeke, T. Coosemans, S. Faid, and J. Van Mierlo, “A data-driven method for energy consumption prediction and energy efficient routing of electric vehicles in real-world conditions,” Energies, vol. 10, no. 5, p. 608, 2017.

[19] H. C. Frey, N. M. Rouphail, and H. Zhai, “Speed-and facility-specific emission estimates for on-road light-duty vehicles on the basis of real-world speed profiles,” Transportation Research Record, vol. 1987, no. 1, pp. 128–137, 2006.

[20] G. M. Fetene, S. Kaplan, S. L. Mabit, A. F. Jensen, and C. G. Prato, "Harnessing big data for estimating the energy consumption and driving range of electric vehicles," Transportation Research Part D: Transport and Environment, vol. 54, pp. 1–11, 2017.

[21] X. Wu, D. Freese, A. Cabrera, and W. A. Kitch, "Electric vehicles' energy consumption measurement and estimation," Transportation Research Part D: Transport and Environment, vol. 34, pp. 52–67, 2015.

[22] S. Amini, I. Gerostathopoulos, and C. Prehofer, "Big data analytics architecture for real-time traffic control," in 2017 5th IEEE international conference on models and technologies for intelligent transportation systems (MT-ITS). IEEE, 2017, pp. 710–715.

[23] E. Husni, G. B. Hertantyo, D. W. Wicaksono, F. C. Hasibuan, A. U. Rahayu, and M. A. Triawan, "Applied internet of things (iot): car monitoring system using ibm bluemix," in 2016 International Seminar on Intelligent Technology and Its Applications (ISITIA). IEEE, 2016, pp. 417–422.

[24] T. Furuichi and K. Yamada, "Next generation of embedded system on cloud computing," Procedia Computer Science, vol. 35, pp. 1605–1614, 12 2014.

[25] S. Teerapittayanon, B. McDanel, and H.-T. Kung, "Distributed deep neural networks over the cloud, the edge and end devices," in 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2017, pp. 328–339