The bases on which autonomous vehicle industry should make this critical decision of choosing between SQL and NoSQL database.

Introduction

Word count:1906 words

With each passing year car companies and other technology companies, manufacturers and researchers of autonomous vehicles(AV) are getting closer and closer to a fully autonomous vehicle in the quest to reduce road accidents deaths and fatalities some leading to long-term disabilities resulting from them as compared to autonomous driving is considered one of the safest technology of transportation as the human error prone factor is eliminated, also, the reduction of traffic and travel time and others. With this in view, major technological industries in developed countries like the United States(US), China and Japan are leading research, though a completely self-driving car dream is far off in the future as trials are still ongoing which results from some vendors not very encouraging, particularly for accident AVs but encouraging in other aspects. Examples of such projects include those pioneered by Nissan and the Google Waymo project and so on(Favarò *et al.,*2017).

Most companies though have made advanced autonomous vehicles and their related technologies. Most of them are not ready yet as the technology is not fully operational and though AV is present on most roads in cities in the US, it is yet to be permitted in Japanese streets except for private streets for even experimentation.(Bianchini *et al.,*2021;Favarò *et al.,*2017;Kellett *et al*.,2019).Despite all the seem bright hope of autonomous, critics point out some problems how far off is the ready technology, who will own these vehicles, and so on(Kellett *et al*.,2019)

In Artificial Intelligence(AI), Computer vision has a two-way goal, biologically to replicate the human vision system and on the engineering, side to replicate and automate tasks that the human nervous system and the sensory system do by getting high insights from videos and images as noted by West, (2016) and J. Bischoff *et al.*(2019).

Autonomous vehicles(self-driving cars) which are one of the many applications of computer vision are one of the new areas in the application of computer vision. It hasn't been fully implemented on the market though there are ample applications in the areas of self-parking cars, and collision and lane departure warning systems to name a few.

Body

The focus of this discussion is to discuss all the bases on which the autonomous vehicle industry should make this critical decision of choosing between SQL and NoSQL databases.

The range of autonomous self-driving cars varies from complete autonomous where vehicles rely solely on computer vision feed by radars,sensors-light detection and ranging (LIDAR) and cameras placed at vantage positions in and around the vehicle to pick big data and process it to accomplish feats like braking, parking, interpret signage and avoid obstacle using collision avoidance as well as generally navigate as,according to (Khan and Al-Habsi ,2020);J. Bischoff *et al.*,2019,Taeihag and Lim 2019).Several well-established methods for prediction and analysis of autonomous vehicle data including supervised, unsupervised and semi-supervised learning amongst the mostly used are supervised and reinforced learning and its applications.The use of software like OpenCV and Tensorflow has helped out in the modelling and training of these models.

Of further interest is the kind of databases SQL or NoSQL which support autonomous machines and data types these autonomous vehicles receive, produce, analyse and transmit.

According to Brown (2020), the explosion of big data has created problems and challenges for autonomous vehicles in data collection and management. "The number of connected devices in 2017 was 8.4 billion and will grow to an estimated 20.8 billion in 2020" and it is increasing according to ('Graph databases lie at the heart of $7 trillion self-driving car opportunity, 2017). With just one AV it is expected to generate over 30 terabytes of data a day as defined by Brown (2020). But as the saying goes, when there is a problem, a solution arises. Modern NoSQL databases provide opportunities to scale with such increasing data sizes.NoSQL can accommodate due to the variety of data models such as key-value, and document, with graph databases like

Neo4j, Infinite Graph and Sparksee as examples leading the pack, hence they are better suited than SQL databases as the NoSQL for storing graphic information like networks of data according to (‘NoSQL Databases List by Hosting Data - Updated 2022’, 2022). Hence graph databases are key in the self-driving car industry. With the choice of NoSQL, there is a need to make use of polyglot persistence-where different storage/stores support various data types and storage needs, hence basically using more than one core database or storage technology as defined by Brunskill (2019)
NoSQL databases are built for data insertion without a predefined schema making it easy for changes in real-time without interruptions leading to less need for a database administrator compared with SQL databases or relational databases with schema constraints, in addition, relational databases were not built for rapidly changing structured,semi-structured and unstructured data and are expensive to run. It can only scale in one direction-monolithic architecture. (‘Microservices.io’, 2017) which is also expensive to run as noted by Brown (2020).NoSQL databases can store, and utilise structures, and semi-structured and unstructured data necessary for multimedia storage needs which are the main types of data that autonomous data process and receive.

Graph database consists of many nodes interconnected ,directed by connections and allows for storage of the links between each entity like the human brain. They are designed for easy,flexible modelling coupled with high performance. They provide support for geospatial data and geolocation data that a self-driving car generates and makes use of. Trying to accomplish this relational databases will be too complex as a lot of joints will be created as well as several new tables too (‘Product’, 2022), Brown (2020)
Examples of datasets of AVs compiled by The KITTI dataset have been recorded from a moving platform while it was driving in and around Karlsruhe, Germany. It includes camera images, laser scans,high-precision GPS measurements and IMU accelerations from a combined GPS/IMU system according to Geiger *et al*.(2013).
They also meet the real and near real time needs of autonomous vehicles normally without requiring large hardware requests.

Like the vision system AV turns to mimic or copy, the human visual system performs these “four cyclic steps: perception, scene generation, planning, and action. The human driver takes in the traffic information through all his senses, especially the vision and creates a 3D environment of the driving scenario around his/her car. Based on this, the driver then plans the sets of actions and finally executes the action through steering, accelerator, brakes, indicators, and so on. The conscious cycle repeats itself throughout driving. A self-driving software prototype can be built exactly along the same principles. It should be able to perceive the driving environment, generate the 3D scene, plan, and execute the action in rapidly iterating cycles" according to Bianchini *et al.*(2021).Making use of supervised and reinforced learning models, the drive process is achieved.

To manage, process and hassle their stream of big data by autonomous vehicle is through a platform, which utilises big data technologies ecosystem Hadoop distributed file system(HDFS)-for efficient and effective data storage such as raw image data, LiDAR data from sensors and its related technologies like, Spark-for big data processing and analysis and Kafka-a messaging platform for real-time messaging between the car and the server(cloud storage) come to the fore which will need cloud solutions for multiple nodes, to reduce data loss due to replication in real across the servers Due to this, there is the need for reliable internet connectivity and backup services. Though an SQL database like MariaDB is utilised for storing data such as “GPS and IMU data, which are sensor data, vehicle driving information, object type identified by Vision, and weather data”, due to the huge amount of data produced it is transferred to the HDFS periodically. (‘Apache Hadoop’, 2022; ‘Apache Kafka’, 2022; Yoo *et al*.2020; ‘Apache Spark - Unified Engine for large-scale data analytics, 2018)

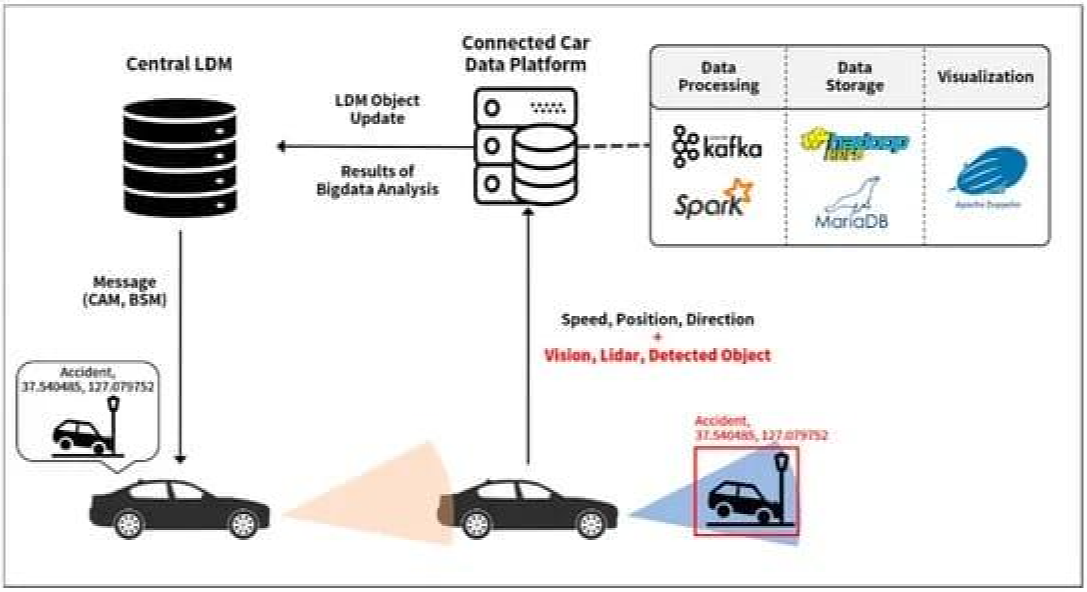

Figure1.1 Sensor big data processing system for autonomous vehicles in the cooperative intelligent transport system environment(Yoo *et al*.2020)

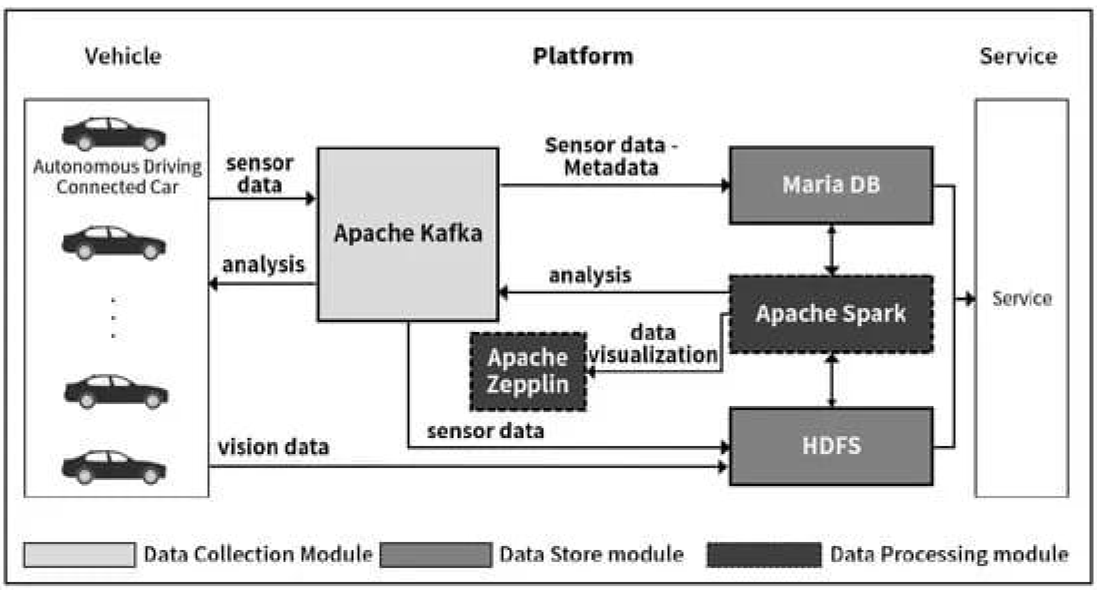

Figure 1.2.Architecture of the platform in the proposed C-ITS environment(Yoo *et al.*,2020)

Other ones include The Waymo Open Dataset, and nuScenes which are similar in size and structure and both provide labelled 3D bounding boxes. The Oxford Robot Car dataset focuses on localization and mapping rather than 3D object detection and semantic segmentation.Further data types, their characteristics and their sources can be found at (Bianchini *et al.,*(2021).

Name	**Size**	**Scenes**	**Map**
KITTI	6 h	50	None
Oxford Robot Car	1000 km	100 +	None
Waymo Open Dataset	10 h	1000	None
ApolloScape Scene Parsing	2 h	NA	None
Argoverse 3D Tracking v1.1	1 h	113	Lane centre lines, lane connectivity
Lyft Perception Dataset	2.5 h	336	Rasterized road geometry

Table 1.3 A summary comparison among the various up-to-date AV datasets according to Bianchini *et al.,*(2021)

The autonomous vehicle is quietly expensive as it relies on expensive gadgets and equipment like highly specialised infrastructure and equipment like Light Detection and Ranging (LIDAR) for navigation, Global Positioning for localization, and Laser Range Finder for obstacle detection and others and on, also automakers want to drop the LIDARs technology as it is bulky according to (Bianchini *et al.*2021; Berrada and Leurent ,2017)

To finally touch on the privacy and security issues.With the increase of adoption of autonomous vehicles,privacy and security will be an issue.In a recent study(Boone,Dave and Roy,2021) enough has not been done in the way of ensuring that the collection, use, choice, and security of consumer data gathered from these autonomous vehicles or autonomous technology is safe enough to prevent the unwanted third-party access of that data as the risk of cyber threat increases around the world. Hackers could potentially access the personal data of a driver, vehicle's location, the identity of others in the car, and whether the driver is home at any particular time. Additionally, cyber -attacks could have potentially fatal consequences, not just for the driver and passengers inside the vehicle, but for anyone or anything physically surrounding the self-driving car as it picks up more data.To reduce this,cloud computing incorporated with RSA algorithm and ORAM should be the focus.Also the adoption of a common open data vehicular database to be

shared amongst multiple collaborators and the concept of concept of Haystack Privacy where-”each data owner
independently and individually privatises their data before privately writing to a cloud service which aggregates all the response and which strengths in privacy strength as more data owners participate yet maintains accuracy,according to in a recent study by (Gerla and Joy,2017)

Conclusion

In conclusion, though autonomous looks to be the next paradigm shift in vehicular technology, according to Brown (2020), the open architecture of NoSQL databases with the integration of polyglot persistence which is scalable, highly flexible and high performance makes it an ideal choice over SQL databases as they can cope with big data ,which can be managed and processed by reliable technologies like the Apache ecosystem of hadoop,Spark and Kafka working in real or near real-time whiles incorporating secure cloud technologies.

References

‘Apache Hadoop’ (2022) Apache.org.2022 [online]. Available from: https://hadoop.apache.org/ [Accessed 1 August 2022].

‘Apache Kafka’ (2022) Apache Kafka.2022 [online]. Available from: https://kafka.apache.org/ [Accessed 1 August 2022].

‘Apache Spark - Unified Engine for large-scale data analytics’ (2018) Apache.org.2018 [online]. Available from: https://spark.apache.org/ [Accessed 1 August 2022].

Berrada, J. and Leurent, F., 2017. Modeling transportation systems involving autonomous vehicles: a state of the art. Transportation Research Procedia, 27, pp.215-221.

Bianchini, M.B., Ghosh, A.G., Balas, V.B. and Shaw, R.S. (2021) Artificial Intelligence For Future Generation Robotics [online]. 1st ed. : .

Bianchini, M.B., Ghosh, A.G., Balas, V.B. and Shaw, R.S. (2021) Artificial Intelligence For Future Generation Robotics [online]. 1st ed. : . ,p.122

Bischoff, J., Maciejewski, M., Schlenther, T. & Nagel, K. (2019) Autonomous Vehicles and their Impact on Parking Search. IEEE Intelligent Transportation Systems Magazine. 11 (4), pp. 19–27.

Brown, K.W., 2020. Using Polygot Persistence with NoSQL Databases for Streaming Multimedia, Sensor, and Messaging Services in Autonomous Vehicles (No. 2020-01-0942). SAE Technical Paper.

Brunskill, V.-L. (2019) polyglot persistence SearchAppArchitecture.2019 [online]. TechTarget. Available from: https://www.techtarget.com/searchapparchitecture/definition/polyglot-persistence#:~:text=Polyglot%20persistence%20is%20an%20enterprise,database%20(DB)%2Fstorage%20technology. [Accessed 24 July 2022].

Dave, R., Boone, E.S. and Roy, K., 2019. Efficient data privacy and security in autonomous cars. Journal of Computer Sciences and Applications, 7(1), pp.31-36.

Khan, A.I. and Al-Habsi, S., 2020. Machine learning in computer vision. Procedia Computer Science, 167, pp.1444-1451.

Geiger, A., Lenz, P., Stiller, C. & Urtasun, R. (2013) Vision meets robotics: The KITTI dataset [online]. The International Journal of Robotics Research. 32 (11), pp. 1231–1237.

‘Graph databases lie at the heart of $7 trillion self-driving car opportunity’ (2017) Information Age.September 2017 [online]. Available from:

https://www.information-age.com/graph-databases-heart-self-driving-car-opportunity-123468309/ [Accessed 24 July 2022].

Favarò, F.M., Nader, N., Eurich, S.O., Tripp, M. & Varadaraju, N. (2017) Examining accident reports involving autonomous vehicles in California [online]. PLOS ONE. 12 (9), pp. E0184952.

Joy, J. & Gerla, M. (2017) Internet of Vehicles and Autonomous Connected Car - Privacy and Security Issues. In: 2017 p. doi:10.1109/icccn.2017.8038391.

Masello, L., Sheehan, B., Murphy, F., Castignani, G., Mcdonnell, K. & Ryan, C. (2022) From Traditional to Autonomous Vehicles: A Systematic Review of Data Availability. Transportation Research Record: Journal of the Transportation Research Board. 2676 (4), pp. 161–193.

'Microservices.io' (2017) microservices.io.2017 [online]. Chris Richardson. Available from: https://microservices.io/patterns/monolithic.html [Accessed 25 July 2022].

'Product' (2022) Neo4j Graph Data Platform.12 April 2022 [online]. Available from: https://neo4j.com/product/#overview%202019 [Accessed 25 July 2022].

Reddig, K., Dikunow, B. and Krzykowska, K., 2018. Proposal of big data route selection methods for autonomous vehicles. Internet Technology Letters, 1(5), p.e36.

Sebe, N., Cohen, I., Garg, A. and Huang, T.S., 2005. Machine learning in computer vision (Vol. 29). Springer Science & Business Media.

Taeihagh, A. & Lim, H.S.M. (2019) Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks [online]. Transport Reviews. 39 (1), pp. 103–128.

Yoo, A., Shin, S., Lee, J. & Moon, C. (2020) Implementation of a Sensor Big Data Processing System for Autonomous Vehicles in the C-ITS Environment. Applied Sciences. 10 (21), pp. 7858.

Yoo, A., Shin, S., Lee, J. & Moon, C. (2020) Implementation of a Sensor Big Data Processing System for Autonomous Vehicles in the C-ITS Environment. Applied Sciences. 10 (21), page 8 &13

West, D.M., 2016. Moving forward: self-driving vehicles in China, Europe, Japan, Korea, and the United States. Center for Technology Innovation at Brookings: Washington, DC, USA.