To help the research company Automobile, we will use different technologies related to Big Data such as Predictive Analytics, NoSQL Databases, Hadoop ecosystem and Blockchain that cover several aspects to complete the company's objective. But for all this, we need to understand very well how Big Data works and how each theme is compatible to solve this.

Big Data is a term that is increasingly important in information technology (IT), relating to other areas such as marketing, business, medicine, governance, among others. This phenomenon refers to the use of a large amount of information with the ability to resist the collection, storage, management and analysis of millions of data, all with the aim of understanding unimaginable knowledge that is not available with conventional analysis tools. With the knowledge gained, we can understand how the behaviour of fuel economy was in 2015 between the different car manufactures at high speed and represented with data analysis models.

It is estimated that around 3 billion terabytes of daily information are generated in the world, such as text, documents, images, videos and music through social networks, applications, and automatic machines, among all these. The amount of data generated in recent years has been growing exponentially and will continue at the same pace for many more.

To have a Big Data visualization, it is necessary to know several pillars known as the analysis of structured data that are necessary to understand the method of analysis that uses Big Data including existing methods that are shared with conventional analysis but with much more data. The administration of the databases is fundamental for the storage and management of the operative data is a very important resource for the Big Data. Object-oriented programming is used for the management and processing of various types of data. The administration of servers is essential because it supports the entire infrastructure of the massive data.

The Big Data process is a composite of four stages from the data source. The first is the acquisition of data that are tools to extract data from its sources. Storage follows, being the platforms that store the extracted and processed data. The next one is the analysis where we find the patterns, relationships and the prediction of the processed data. Finally, the visualization refers to the report of all the work to inform in the clear way in which the knowledge is obtained.

For Big Data, data management should consider the following characteristics: The volume that refers to the amount of data; a variety that includes all types of structured and unstructured data, and as third the speed when making decisions in almost real time. There are two types of Big Data applications, the first is to improve operations and detect deviations, the second focuses on business decisions based on the complex analysis. All this to reduce the cost and time, the implementation of personalized offers and business intelligence.

The Big Data ecosystem is managed by the information technology (IT) system in two parts: The first is the administration and storage of data that are divided into three areas: source data (structured and unstructured), Big Data core that contains the data along with management tools such as ETL and the architecture that processes the distributed data; and the administrator of operational data that contains the data processing. The second part is the analysis and applications divided into two areas: Big Analytics, which is a division in the development environment and the analysis products, and users as business analysts.

The data flow in Big Data is as follows: The data is obtained, the unstructured data will be processed in the kernel with MapReduce and the structured data will be sent to the administration tools such as ETL. The data that comes out of the kernel is NoSQL data and will become operational data that will help the Big Analytics section feed the business analysis and applications.

Big data requires several elements that work with each other. The flow of the life cycle happens to create the processes that will take advantage of the big data; the model is composed of fourth micro processes, and only is possible if there is a complete identification of the data:

1. Record and recollection of data in several sources
2. Filter, enrichment and classification of data
3. Analysis, modeling and prediction of data
4. Delivery and visualization of data

The infrastructure of the big data is through clusters that are a set of independent computers but connected between them. Each computer is called node have the function as supercomputers its advantages are:

* Parallel work provides fault tolerance with continuous availability.
* It has a high-performance capability for the operation of multiprocessors.
* Support of high workloads because it is distributed among all the available nodes.
* Great scalability because more hardware can be added if necessary
* The software of the clusters is mainly an open source that works as a distributed file system.

There are a lot of technologies that can help us in our car research, but for the moment we will only name four of them.

**Predictive Analytics:** It is a technology to predict future events based on the performance and behaviors compiled from a historical object to analyze, which will help business decisions. This technology can help us to influence the audience interested in the automotive world when we want to work with marketing plans as a statistical base as well as the generation of a new audience. We can also know what the future needs of a customer sector are in terms of products and services related to automobiles, and can be covered by having this knowledge in advance.

You can also know how far the company can grow based on the influence of customers to make strategic plans and reduce risks and costs, minimizing uncertainty by providing enough information to make quick and specific decisions. In the same way, with predictive analysis, we can calculate an estimated price that each product or service can have, helping to accelerate innovation and outperform the competition.

**NoSQL databases:** It is a new way of storing unstructured data such as images, sounds and videos, in addition to having the capacity to store large volumes of data without affecting its performance because it can be scaled horizontally by adding more nodes to the system. It does not implement rigid methods of consistency so any type of scheme can be added to the structure, adding the new fields that are necessary for the investigation without the other information being affected by this, so if we want to add a new vehicle manufacturer that has different characteristics to those already worked, we can add it without any concern and in the same way work with those data in the analyzes that are needed.

In addition to the NoSQL database systems have a speed-oriented paradigm so that make storage and search for information every time is faster taking better advantage of storage resources such as hard drives, because it will only store information necessary.

**Hadoop ecosystem:** It is free software that allows the handling and processing of large volumes of information. For example, the information collected from the cars will be handled at high speed because it is distributed among the different nodes giving each one power to the system, besides that there is no need to worry about if some node goes to fail and it gets to lose information, because Hadoop distributes copies throughout the system where another node can continue with the execution of the task, this distribution is through the Map-reduce algorithm.

Another advantage that this technology allows us is that it is not necessary to work with large hardware requirements to start using it, because its scalability allows adding nodes as required by the investigation. Finally, Hadoop has the possibility of receiving data from other platforms such as Data Warehousing or relational databases as well as migrating data to other platforms to perform a graphical analysis.

**Blockchain:** It is a technology that allows the encryption of information and the ability to make processes more efficient, transparent and secure. It supports large amounts of data when conducting transactions that also have a high degree of confidence and protection of the data due to the fact that the stored information cannot be altered, nor can everyone see the contents of the transaction, thereby resisting malicious attacks. They can happen by altering the results of the analysis of the research on automobile factories. Another advantage that is had with the blockchain is its easy use, so that intermediaries can be avoided when manipulating private information.

Having the appropriate knowledge in these technologies, it can be affirmed that the use of them will greatly facilitate obtaining the best results required for the company's research, in addition that the execution times will be much lower compared to being performed by manual methods.

**REFERENCES**

Mattturck.com (2018). *Great Power, Great Responsibility: The 2018 Big Data & AI Landscape.* [Online] Available at: [*http://mattturck.com/bigdata2018/*](http://mattturck.com/bigdata2018/) *[Accessed Feb. 2019]*

Whatsthebigdata.com (2017). *Big Data Ecosystem and Benefits.* [Online] Available at: [*https://whatsthebigdata.com/2017/03/28/big-data-ecosystem-and-benefits/*](https://whatsthebigdata.com/2017/03/28/big-data-ecosystem-and-benefits/) *[Accessed Feb. 2019]*

Ibm.com (2018). *Predictive analytics*. [Online] Available at: [*https://www.ibm.com/analytics/predictive-analytics*](https://www.ibm.com/analytics/predictive-analytics) *[Accessed Feb. 2019]*

Sas.com (2018). *Predictive Analytics: What it is and why it matters*. [Online] Available at: [*https://www.sas.com/en\_us/insights/analytics/predictive-analytics.html*](https://www.sas.com/en_us/insights/analytics/predictive-analytics.html) *[Accessed Feb. 2019]*

Mongodb.com (2018). *NoSQL Databases Explained*. [Online] Available at: [*https://www.mongodb.com/nosql-explained*](https://www.mongodb.com/nosql-explained) *[Accessed Feb. 2019]*

searchdatamanagement.techtarget.com (2018). *Hadoop*. [Online] Available at: [*https://searchdatamanagement.techtarget.com/definition/Hadoop*](https://searchdatamanagement.techtarget.com/definition/Hadoop) [Accessed Feb. 2019]

coinsutra.com (2019). *The Ultimate Guide To Understanding What A “Blockchain” Is & How It Works*. [Online] Available at: [*https://coinsutra.com/blockchain/*](https://coinsutra.com/blockchain/)[Accessed Feb. 2019]