All the technologies presented here can help us to manage and solve several problems found in the moment of work with big amounts of data of any origin. For this, it is introduced to each topic the explaining how is the functionality of that technology, which is its relation with the big data and how can help to resolve the problem presented in the Automobile research company.

**Predictive analytics**

This technology helps us to make predictions about unknown future events through of several techniques that permit to try big amounts of data that it can be used to identify opportunities and risks that allow organizations to become proactive anticipating outcomes and behaviors based upon them.

The process for the predictive analytics can be detailed of the following way:

1. Define the project: the results, deliverables and objectives for the project are established, as well as the data sets that will be used.
2. Data collection: through the extraction of data from multiple sources, they are prepared for their analysis, thus discovering patterns, trends and behaviors that will become a comprehensible structure for future uses.
3. Data Analysis: it permits inspecting, cleaning of outliers and other incorrect information, transforming and modeling data with the objective of discovering useful information with diverse techniques for its analysis.
4. Statistics: It uses statistical models that permit the study of the collection, organization, analysis, interpretation, and presentation of data to test them and validate the hypotheses established.
5. Modeling: Here a model is chosen based on the tests, validation and evaluation that can be used by one or more classifiers to determine the probability that one data set belongs to another set.
6. Implementation: the models are validated and then passed to production to implement a scoring system where the model applies new data that does not have a dependent variable.
7. Supervision of the model: this step is necessary to ensure that it provides the expected results.

Predictive analysis can help us make the best and fastest decision regarding fuel economy research in different car manufacturers to create new products or services that customers may need in the near future, as well as anticipate and identify the opportunities that will help the organization.

**NoSQL Databases**

Is a new breed of database management systems that fundamentally differ from relational database systems because don´t require tables very structured. NoSQL databases have big popularity in the management of the data for several reasons:

1. Scalability: offers the option of growing distributed in multiple pre-existing hosts.
2. Reduce the dependence of specialized employees: the maintenance of a Relational Database System is only possible using trained DBAs, involved in the design, installation and performance of these, which certainly are not cheap. NoSQL databases are designed to require less practical administration in data distribution, automatic repair and simplified data models.
3. It's cheaper: NoSQL databases are designed to use server clusters for the management of ever-growing transactions and data volumes.
4. Few restrictions: NoSQL does not have many restrictions in its database models allowing relative flexibility as the addition of new columns without major efforts.

Therefore, with NoSQL databases, we can manage large volumes of data without having to worry about scalability or cost, helping us to store all available information about each car model, which contains 729 entries and 19 attributes of the different car manufacturers. If changes are made to these data, they can be done in a faster and easier way, giving as an advantage the flexibility over structured databases that are not capable of supporting large volumes of data.

**Hadoop ecosystem**

Hadoop is open source software with the power to store, manage and analyze large amounts of structured and unstructured data in a fast, reliable and flexible way to help companies manage all that information that is generated daily. The capabilities that Hadoop has are:

1. Scalability and performance: the distributed file system of local data to each node in a cluster allows to store, manage, process and analyze data on a large scale in large groups of servers that are expanded to hundreds of petabytes.
2. Reliability: large clusters are prone to fail, Hadoop is resistant when a node fails, processing is redirected to the remaining nodes in the cluster and the data is automatically replicated in preparation for future failures.
3. Flexibility: it does not need to create structured schemas before storing data
4. Low-cost: Hadoop is open source software and runs on low-cost hardware.
5. Data processing: Hadoop use originally a framework called MapReduce for writing massively applications that process large amounts of data stored.
6. Data access and analysis: In Hadoop, the applications can interact with the data using batch or interactive SQL or low-latency access with NoSQL to allow business users and data analysts to use their preferred business analytics, reporting, and visualization tools.
7. Data governance and security: the Hadoop ecosystem extends data access and processing with powerful tools for data governance and integration including centralized security administration and data classification tagging, which combined enable dynamic data access policies that proactively prevent data access violations from occurring. Hadoop perimeter security is also available to integrate with existing enterprise security systems and control user access to Hadoop.

With the Hadoop ecosystem, we have the ability to store a large variety and amount of data quickly from any resource and decide how to use it in automotive research at any time. The powerful distributed model works at high speeds, while more nodes are added to the work group because it increases the processing power obtaining the necessary analysis of the cars in a very short time and at low cost.

Fault tolerance is a very important characteristic because if one node fails, the rest of the work is distributed automatically among the other nodes, storing several copies of the data and without worrying that the results of the analysis of the investigation are altered by losses of information.

**Blockchain**

It is a distributed database technology and secured by cryptography to store and send information between nodes. Each record in the database is called a block which contains public and private information, such as transaction data and a link to the previous block.

The operation with blockchain can be divided into several stages:

1. First, the operation begins among the users of the network by joining their data in blocks
2. Each block is validated for the nodes in the network using cryptographic techniques.
3. The already validated block is added to a block chain that cannot be modified and is available to all users.
4. Finally, the system seals and completes the transaction.

In this way, the blockchain offers the following advantages:

1. Decentralization: users make transactions through blockchain, managing to enjoy a Peer to Peer system where they do not depend on any entity (such as banks) to centralize the data and validate their operations.
2. Protection: the blockchain represents a high security technology for high volume information systems and with different types of databases.
3. Versatility: this technology can be used not only for security in cryptocurrencies, but it can protect many types of data.

With the processing of large amounts of information through Big Data, it helps research to make more precise decisions. However, several techniques raise doubts about data protection and unauthorized access to them. For this, Blockchain joins Big Data by offering security where the fraud detection process is guaranteed in real time, making a non-modifiable data record and thus protecting the veracity of the data given about the attributes of each car. When the information is collected, the block chain will allow transparency to manage the data, since it cannot be changed without the permission of the entire server in the network.

As we can detail, each technology here called to help us with the manage, store, security and analysis of big amounts of information with matter the origin of the data and we can use to solve the problem of the Automobile research company.

**REFERENCES**

Predictiveanalyticstoday.com (2014). *What is predictive analytics*. [Online] Available at: [*https://www.predictiveanalyticstoday.com/what-is-predictive-analytics/*](https://www.predictiveanalyticstoday.com/what-is-predictive-analytics/) [Accessed Feb. 2019]

Bigdata-madesimple.com (2015). *Advantages of NoSQL Databases – Everything you need to know*. [Online] Available at: [*https://bigdata-madesimple.com/advantages-of-nosql-databases-what-you-need-to-know/*](https://bigdata-madesimple.com/advantages-of-nosql-databases-what-you-need-to-know/) [Accessed Feb. 2019]

Blogs.harvard.edu (2013). *NoSQL vs Relational DB.* [Online] Available at: [*https://blogs.harvard.edu/acts/2013/06/26/nosql-vs-relational-db/*](https://blogs.harvard.edu/acts/2013/06/26/nosql-vs-relational-db/) [Accessed Feb. 2019]

Sas.com (2018). *What is Hadoop?.* [Online] Available at: *https://www.sas.com/en\_us/insights/big-data/hadoop.html* [Accessed Feb. 2019]

Weusecoins.com (2018). *What is the Blockchain?.* [Online] Available at: [*https://www.weusecoins.com/*](https://www.weusecoins.com/) [Accessed Feb. 2019]