Digitization Process of Test Bench for Automotive Components

Lucas Leonardo Goy
Universidade Tecnológica Federal do Paraná – Ponta Grossa
Robert Bosch GmbH

Willian Keiichi Tesuka
Gustavo Scherpinski
Robert Bosch GmbH

Max Mauro Dias Santos
Universidade Tecnológica Federal do Paraná – Ponta Grossa

ABSTRACT

This paper addresses the creation of data pipelines for the digitization of test bench sensoring. The proposed pipeline covers data acquisition, cleaning, transformation, analysis, and storage.

Acquisition techniques, including compatibility with sensors and real-time monitoring systems, are discussed in this paper. The transformation step involves standardization and integration of data formats. Analysis methods are discussed to extract relevant information. Secure storage and making the results available in dashboards are the main topics covered.

Implementing a data pipeline brings benefits such as efficiency, cost reduction, and informed decision making. This article provides an overview of the key components for digitizing test benches.

INTRODUCTION

The transportation sector is directly linked to the progress of a nation [1], this sector is responsible for managing and distributing all of its resources through its territory. In Brazil, which has a vast territorial extension this topic is presented with even greater relevance given the fact that the demand for transportation is focused on cargo transportation posing a national problem.

According to the CNT (National Transportation Confederation) about 64.7% of all cargo transportation carried out within the country is conducted via highways [2], that is, through the use of trucks or similar vehicles that present themselves as major contributors to the unbridled advance of climate change.

With the increase in restrictions on vehicle emissions and a continuous search for improvement in the performance of internal combustion engines, the need for improvements becomes increasingly greater, demanding an rise on the number of bench tests [3]. Thus, the industry looks with good eyes the constant evolution and updating of the components that are part of the engine aiming at a reduction in gas emissions, as well as a better efficiency.

For this improvement, it is necessary to carry out controlled tests in environments that can be monitored in detail, with a whole structure of sensors capable of measuring the variations according to the demands of the engine or product. For this type of "simulation", industries use test benches, given their ability of monitoring and providing all the necessary data so that the engineer can validate and update the product design.

To this end, it is necessary to introduce the importance that digitization plays in industrial processes today. Observing from a practical side there is a growing demand for data capture and especially data consumption since this wave allows a greater technical basis for the engineer when making decisions about line products.

However, the extraction of information from these large volumes of collected data is not easy, it requires the development and application of tools that enable its visualization and present a trend curve of product operation, highlighting design flaws or adverse functioning in stress situations.

According to Brescovitt [4] the digitization process has been gaining strength within industrial processes in view of the reduction of labor and its increased complexity. In addition to this, the author mentions that the traceability of the data generated by the sensoring of the process allows an information curve with a degree of definition in the millisecond range, generating greater reliability in the tests conducted in laboratory, which were previously carried out through follow-ups by hand-filled spreadsheets. In addition to this, the storage of these results can generate future fruits when compared with other trend curves of tests previously performed under similar application conditions.

With a focus on the automotive industry, which is the field of application of this work, the digitization of processes has been gaining enormous prominence, due to the high visibility that this type of project generates given the requirements raised by EURO 6 in Europe and in Brazil by PROCONVE, standards that impose an increasing responsibility on automakers to reduce CO_2 emissions.

TEST BENCH

For this article it is necessary to contextualize what constitutes a test bench, what is its function within the industry and its operating mode. Figure 1 shows a P-Diagram of one of the benches used in this work as pilot bench for the development of the pipelines.

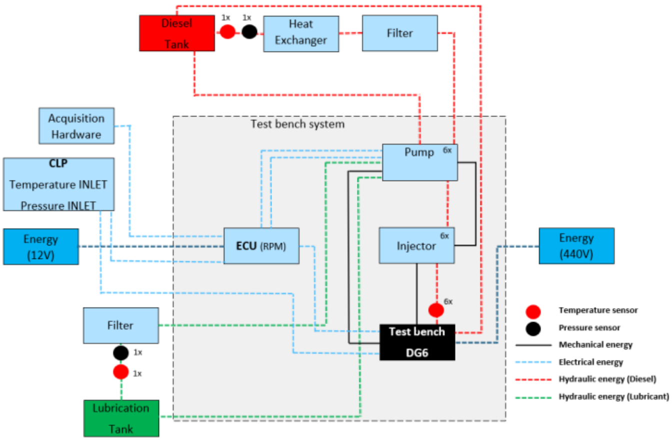

Figure 1. Physical Structure Bench D08

Explaining the operation of the bench used to conduct the pilot project of the durability laboratory, we have that:

1. The bench consists of two main parts, the mechanical part represented (in black) in the image constituting the operating casing of the equipment. In this case we will compare its operation as that of a diesel engine for the sake of simplification.

2. The second part consists of the electro-electronic components (topics in light blue) which are composed of the sensors that indicate to the engineer or technician responsible how the product is behaving as the test time elapses, as well as the data collection of the sensors and control of the bench, in this case represented by an ECU, and a collection module called Acquisition Hardware.

With the basic structure presented, its operation resembles a diesel engine with changes ensuring that the company's products are evaluated in the best conceivable way. The bench in question has compatibility with two company line products (part represented by the Pump and Injection blocks), which makes the need to improve the test bench much more relevant.

Finally, its operation is under instructions and calibrations imposed on the ECU by the hardware acquisition. Being the ECU responsible for controlling the test time and other test variables.

PROBLEMATIC

The focus of this application was to develop two practical models for data handling. The first model arose from the demand for durability tests (bench evidenced above) that can require a load of up to one thousand hours of bench running. This forces the equipment to run full time for prolonged periods, being stopped only for eventual maintenance and calibration. For this application, it was necessary to develop a pipeline structure capable of transmitting the test monitoring data near real time, providing readable information from the sensors present on the bench automatically and remotely.

For the second model that arises from the demand of performance laboratories, the objective was to develop a pipeline capable of processing the data after the tests are complete. This new model permits the engineer to visualize the final curves of the experiments, generating a model with a focus on processing according to the chosen test model, varying between Shot to Shot, Injection Rate, Injection Map among others that may be requested to evaluate the product.

DATA STREAMING MODEL

Nowadays, due to the constant advancement of technologies and the high investment necessary to change equipment, updating the benches is imperative to keep them competitive and in accordance to testing demands required by the industry.

With the theme of digitization and automation in focus by the company, it was found that the durability benches did not have any type of connectivity with the company's network, which forced the work of transferring and collecting data to be conducted manually by one of the laboratory technicians or engineers.

However, the connection of the benches to the network directly does not meet the company's rules regarding data privacy and security, due to the data generated on the bench being of a confidential nature, in addition to the fact that the equipment has operating systems that do not undergo constant updates, being vulnerable to viruses or cyber-attacks. Therefore, it is necessary to stipulate routes for the transit of safe and efficient information that can meet the demands of the laboratory team, without compromising the security of data and equipment.

With the due pains of the laboratory team mapped, an initial study was conducted to verify the feasibility of developing the project.

For an initial project five benches would be within the company's performance map, considering that an engineer needs 15 minutes a day to check the status and bench sensors and that two of the five benches are running at the same time every day of the year, there is a saving of 125 hours of the engineer. In addition, at the end of the test four hours are required to prepare the test report, if twenty tests are considered completed per year, we have another 80 hours saved, totaling 205 hours saved per year. This way, the development of the pipeline can generate a direct financial return for the company.

Analyzing non-measurable gains, there is an increase in test reliability, greater capability of accurate point curve monitoring, preliminary data analysis, possibility of interrupting the test earlier if sufficient results have already been obtained, in addition to the downtime that can be avoided by monitoring the sensors remotely that cannot be measured.

Considering only the costs of engineer hours that this application would save per year, it was determined that this demand was extremely important for the organization and that the project should continue.

Thus, the project was divided into three development fronts, with a single condition in common: the need to develop a standard pipeline model, capable of being replicated for all laboratory equipment with low effort. The project fronts will be presented below.

1. Data Acquisition - This stage consists of capturing all test sensors, in a more general view. This stage presents the greatest inconsistency, since each bench has a set of unique and specific sensors for its tests. Thus, the biggest difficulty for this stage was to find a solution that presented a flexibility of inputs capable of meeting all the demands of the laboratory.

2. Data Streaming - Responsible for all the information security part to be able to carry out the transportation of data between the two ends of the pipeline without interference, undue access or communication errors.

3. Data Visualization - This stage is responsible for creating dashboards to consume the data generated by the bench in real time, the step that needs to be as standardized as possible, because it is in direct contact with the end user of the application.

With the project fronts defined, a bench was selected to be adopted as a pilot project for the elaboration of the entire pipeline. This bench, which will be called Bench 08 and has already been presented before, is composed of eight thermocouples of different models and two pressure sensors, the tests performed on it require many hours, so as this requirement is applicable to other equipment, the need to develop a robust and fail-safe system, considering that the malfunction of the pipeline during the test can generate great losses.

With this in mind, it was researched a hardware model capable not only of performing data collection from the sensors but also of efficient communication protocol to integrate with the next step that consists of data streaming.

After a through performance research, it was found a hardware of the company itself with great compatibility with the demand required in the project, in addition to presenting its structure based on Linux and a modular system capable of performing the collection of information both from the sensors directly by the I/O system ideal for the initial application, as well as having compatibility with PLC's and other devices.

This model has compatibility with the most used communication protocols within the industry for data streaming which are the OPC-UA, MQTT and Kafka protocols, which opens a range of possibilities of structures for the development of the data streaming work front.

The hardware in question can read the sensors much faster than required by the project. Bearing in mind that because it is such a long test, pressure, and temperature variations in the range of seconds are not so interesting when we extend this collection for months, due to the large volume of data that can be generated.

The chosen model has a plug and play structure, also an extremely high processing capacity, so following the indications of the hardware manufacturers, it becomes possible to use only one module for the entire laboratory, thus making the solution cheaper, diluting the value between the five equipment. In the future, this application can be extended without generating great costs for other benches outside the initial package.

For company security reasons, the benches are in a network structure and cannot have direct access to the internet to avoid data leakage due to external attacks. The second stage of the project is then responsible for consuming the data and bridging the gap between it and the visualization system inserted in a network layer with free internet access for data consumption by internal users of the company.

The chosen protocol that fits the rules of the company and that is ideal for the consumption of data in large volume which is the case of data generated in near real time was the OPC-UA protocol, which consists of a data protocol created specifically for data monitoring, with this model the data can be overwritten if there is a variation in the measurement. This ideal for the application since it decreases the volume of data generated. Through this communication model it is possible to leave the 'old' model that consisted of manually collecting data on site, migrating to a much more complete and robust data collection model.

For the communication it was chosen the data streaming model Apache Kafka, a scalable, fault-tolerant, and distributed streaming platform that can be used for building real-time data pipelines and streaming applications which follows publisher-subscriber mode [8], model that through the addition of a buffer, store data for a certain period of time, in order to make it possible to later generate final reports with the entire sensor curve.

For the consumption of data on the other side of the Kafka protocol there is a Python application developed internally by the company that stores data in InfluxDB so that they can then be consumed and exported to the last step named Data Acquisition. Through real-time visualization software, dashboards are created for data visualization by the end user.

The advantage of using online data monitoring software is that with the software interface itself it is possible to create different models of visualization of the data so that when the readings taken by the sensors leave a certain operating range the system issues an alert warning, warning the operators that the test could be in trouble or under other negative circumstances.

With this, the project aims to drastically reduce the idle time of the benches, adding value to the test for the engineer to conduct his product reports. presenting a much more detailed curve, greater resolution allowing to identify the behavior of the bench, anticipating failures, and mapping imminent problems only by analyzing the reading curves of the sensors.

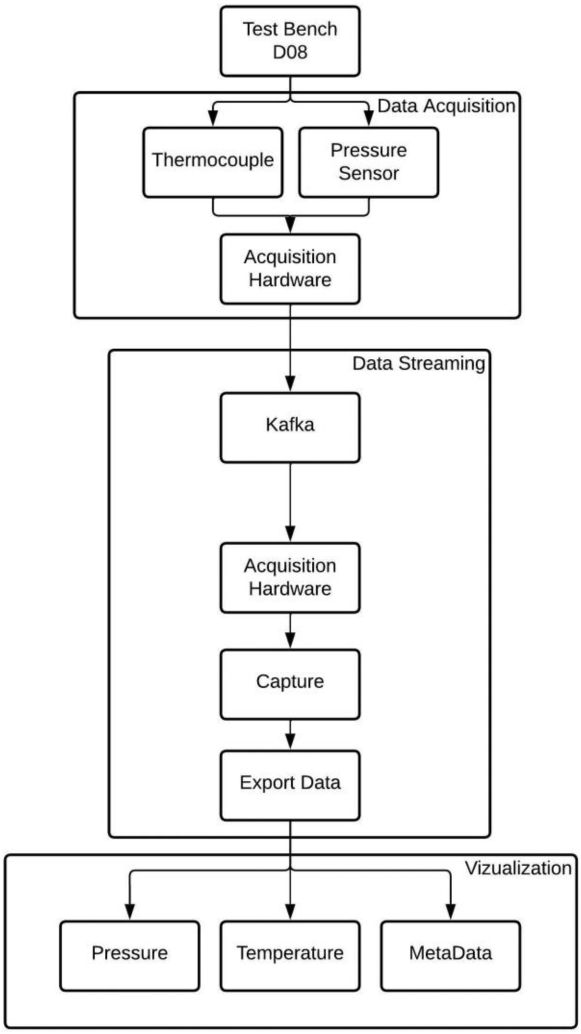

Figure 2. Bench 08 Pipeline Structure

The figure above shows a summarized version of the structure used to process the data generated by the test bench described in this topic. The elaborated system presents a simplicity of understanding which makes the elaboration of improvements in the pipeline much more subtle.

FILE PROCESSING PIPELINE

With the development of a pipeline for real-time data monitoring, a new demand arises within the laboratory, now focused on another area, the performance sector.

Unlike the durability benches that spend exhaustive hours to perform the tests, for the performance benches we have a much more extreme operating system, that is, that forces the product in extremely aggressive conditions in a brief period.

For this equipment the goal is to take the products to the limit in order to follow their behavior in the μs range, something that for durability tests is not necessary since the changes are not significant in such a short period of time.

These benches can perform many more tests in a shorter interval and at the end of each test a file is generated with all the measurements taken. These measurements can range from current profile, to injected quantity measurement among many other test variables.

In this perspective the need arises differently for these benches since it is now much more interesting to wait for the tests to finish and then process the results.

For this pipeline model, once again the project was divided into three major development steps:

1. Data Acquisition - Unlike the previous model this step is not so tied to the data capture itself, because in these benches the files are already generated automatically by the bench. In this model, the data acquisition step is responsible for standardizing the data coming out of the benches.

   Due to the great diversity of sensors present in the performance laboratory and due to the origin of this equipment being different, mostly imported from other plants. There is no standard file format for the benches.

   Therefore, this step aims to develop a model capable of standardizing these input files, in order to create a solution that can receive several formats without presenting possible compatibility problems.

2. Data Preprocessing - Responsible for the entire file processing part, in this step the file is optimized for BigData, this is due to the fact that performance tests need to be stored forever for future comparisons with previous results.

3. Data Analysis and Visualization - This presents itself as the most important part of the entire pipeline, in this model the monitoring of the database and all the processing of the labels of the files according to each test requirement is conducted in this step, in addition to all the elaboration of the visualization dashboards.

With the project fronts defined, a test bench commonly called FCR was chosen as the base project.

Its operation is due to a completely different way from the bench explained above. However, as the pipeline in this case starts after the file generation and not when reading the sensors as in the previous model, for this application it becomes much more interesting to explain the format of the test itself than the operating model.

The file generated by the bench consists of four test steps, namely:

1. VL – *Voll Last;*

2. EM – *Emission;*

3. LL – *Lierlauf;*

4. VE – *Voreinspritzung;*

Each of them forces the products to operate in different ways which generates for the engineer a more dynamic analysis of the product operation.

Starting with the Data Acquisition stage, the main challenge was to develop an application capable of supporting different file inputs in different formats and standardize all of them into a standard data model that is easily compatible with the other tools in the framework.

No tool was found on the market today that had this wide compatibility for file transformation, so the way out was to develop a local solution for the project needs.

This bench presents a very non-standard file structure, because for the tests performed four .txt files are generated for each set of four products, so in a standard test we have eight products generating eight tests.

The bench has four positions allowing the test of four products at once, however the test requires an extremely high precision, so for the engineer to be able to conduct their report it is necessary to redo the test again, until all four components have passed through all positions.

For a better understanding of the structure that is presented in an important way to understand the first steps of the pipeline model presented for this problem, the following table is presented:

File	Pump			
1	A	B	C	D
2	D	A	B	C
3	C	D	A	B
4	B	C	D	A
5	E	F	G	H
6	H	E	F	G
7	G	H	E	F
8	F	G	H	E

Table 1. Structure of the files

Considering that each letter represents a product that is being tested and that all products need to pass through all positions we have at the end a set of eight files.

With the test dynamics explained we have the first challenge, in addition to the eight files generated in .txt, a parallel file is also generated in .xlsx format with some fixed measurements used to perform the visualizations.

Due to the complexity of the demands of the project and its specific condition, it was necessary to develop an application in Python capable of reading and extracting the information from these nine files. In order to save in a .csv format so that it is possible to continue the pipeline.

This step of the process occurs right after the bench finishes generating the test files. With the files properly formatted by the local application, this file is saved in another folder monitored by an application developed in Czech Republic's Bosch Filial, which runs locally on a Workstation. This tool transforms them to .mf4 format accepted for the other steps of the process and communicates with the test control system so that it is possible to attach this data extracted from the test with the meta data taken from the system.

In the Data Preprocessing stage, the entire process of data transmission through the test control system takes place until it reaches the place where the files will eventually be stored.

For this the local application forwards to the test monitoring system a search address for an API system, when the test monitoring system receives this information it communicates with the system and transfers these files from the local application to another application that runs on a server in Germany that captures this file in .mf4 and transforms it to .parquet which is a much more friendly format for large volumes of data storage and saves them in HDFS spaces [5] which is a popular open source implementation of the MapReduce programming model for dealing with large data sets.

These spaces are organized by products, i.e., this way you have all the test information in one place, so that different benches can store data in the same space.

Finally, there is the data analysis and visualization step, considering that the previous step works on the structure of the file to store it in the best possible way, in this step the content of the file becomes the focus.

For the initial bench only one file model is generated because the bench was developed only for one application, but in the vast majority of benches several types of tests are performed so that the structure and the way of reading the results must be built differently.

For this, the pipeline model built presents a pattern defined by a master called orchestration pipeline that is responsible for controlling the frequency of operation and organizing all the others, there are also numerous sub pipelines one for each test performed on the bench, which captures the specific information of each test structure.

In this way it is possible to connect this storage space to the desired visualization tool. For this model tools that have a greater capacity for data transformation and processing are much more interesting, so the range of software available on the market that meet the demands are enormous, leaving the choice to the engineers responsible for the bench to choose by affinity.

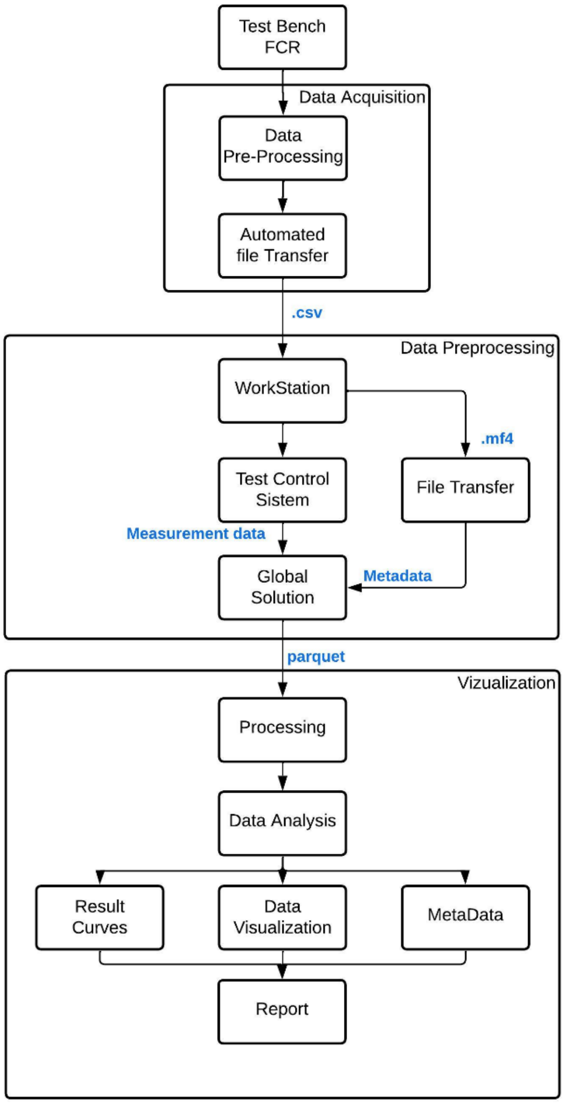

Figure 3 - FCR Structure

In the figure presented it is possible to observe the summary structure developed for the bench models with file generation, this implies that the data pipeline monitors the bench folder so that every time a new file is generated the system identifies this file and forwards the results to a database so that the engineer can interact through the user interface and make it possible to visualize the test results, in a clean and interactive way, in order to meet all possible demands.

CONCLUSION

This work aimed to develop two data pipeline models capable of supporting the existing testing demands within research laboratories.

The standardization of models for data collection, processing and visualization generated several gains evidenced in the table below:

Topic	Old Model	Data Pipeline
Data Collect Model	Data collection was conducted locally by an operator in a spreadsheet	Automated collection
Data Collection Frequency	Performed once a day	Every one minute or less according to test demand
Stop Detection	When an operator walked by the equipment or during daily data collection	As soon as the bench performs a non-standard measurement, the operators are alerted, even before the equipment stops
Traceability	As the control was conducted by a manual spreadsheet, it was necessary to transfer this spreadsheet to an excel and save it in a folder for future consultations	Data is saved so that future reference to any bench test can be performed for comparison with more recent tests.
Visualization	It was necessary to wait for the tests to end and only then could an operator export the data to an excel and plot the results	Visualization of measurements in near real time

Table 2. Comparison of earnings

Among all the factors mentioned above, it is possible to highlight two major gains, which were:

In terms of security, it was that the two structure models were developed within the standards requested by the company, data flow is subject to strong encryption so that the benches do not have direct access to the internet network and in the opposite direction users do not have access to the benches by the data processing model thus fitting the security standards.

In the financial topic now both with the collection of sensors being performed and presented digitally, as well as the visualization of the data already processed after the end of the tests, ends up generating a return of machine and labor that exceeds R$ 50,000.00 annually.

Building an efficient and robust data pipeline is critical to the success of data analysis and processing projects. Throughout this article, the main components of a data pipeline have been explored, from the collection and ingestion of data into the information flow to the storage, processing, and visualization of results.

During the process of building a data pipeline, it is important to consider aspects such as data standardization, the choice of appropriate tools and technologies, the implementation of good security practices and the scalability of the system.

It is important to note that this article has only presented an overview of the process used to build a data pipeline. There are several approaches, tools and techniques that were not addressed in depth due to the protection of the company's intellectual property, and each project may have specific requirements that demand customized solutions.

REFERENCES

[1] COLAVITE, Alessandro Serrano; KONISHI, Fabio. A matriz do transporte no Brasil: uma análise comparativa para a competitividade. Simpósio de Excelência em Gestão e Tecnologia, v. 12, p. 28, 2015.

[2] CNT. Transporte rodoviário de cargas 2021. Confederação Nacional do Transporte. Availabre in:<www.cnt.gov.br >. Access in: 17 mai. 2023.

[3] CEZAR, Gustavo Vianna. Desenvolvimento de uma bancada de testes para motores a combustão interna. Pontificia Universidade Católica do Rio de Janeiro, Rio de Janeiro, agoste de, 2012.

[4] BRESCOVITT, Alexandre Henrique et al. Desenvolvimento de software para integração de instrumentos e aquisição de dados em bancada de testes de motores de indução. 2020.

[5] Tom White. Hadoop: The Definitive Guide. O'Reilly Media,2012.

[6] Smith, J. (2020). Building Data Pipelines: A Hands-on Guide for Managers and Engineers. O'Reilly Media.

[7] Marz, N., & Warren, J. (2015). Big Data: Principles and best practices of scalable real-time data systems. Manning Publications.

[8] Chowdhury, Arnab Ghosh, et al. "An approach for data pipeline with distributed query engine for industrial

applications." 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). Vol. 1. IEEE, 2020.