AUTOMATING DATA COLLECTION AND ANALYSIS

Project Initiation Document

Projectcode:

Date of completion: <Datum voltooid>
Author: Burak Arslan
Student number: 3368246

Version: 0.2
Status: In progress

Document ID:
Bestandsnaam: Project initiation document

Fontys HOGESCHOOL ICT

Document history

Revisions

Version	Status	Date	Made changes
0.1	In progress	17-9-2023	Chapter 1 to 3 added
0.2	In progress	5-10-2023	Chapter 4 to 9 and glossary added,

Approval

This document carries the following approvals:

Version	Date	Name	Function	Initials

Distribution

This document has been sent to:

Version	Date	Name	Functions

Referred documents

This document contains referrals to the following documents:

Document	Title	Date	Version	Author

Table of contents

1 Glossary

Term	Definition
Agile	A set of principles and values for software development that emphasizes customer satisfaction, collaboration, flexibility, simplicity, and quality. Agile methods aim to deliver working software frequently and respond to changing requirements and feedback.
CanBUS	A Controller Area Network (CAN) bus is a vehicle bus standard that allows microcontrollers and devices to communicate with each other without a host computer. It is used for multiplex electrical wiring in automobiles and other applications.
Can LOGGER	A device or software that records and stores data from a CAN bus, which is a communication system used in vehicles and machines.
Chronograf	A user interface and administrative component of the InfluxDB 1.x platform. It allows users to build dashboards, query data, manage tasks, and configure alerts for time series data stored in InfluxDB.
DTC file	A file that describes the structure and format of the data in a trace file.
Grafana	An open source analytics and monitoring solution for every database. It allows users to query, visualize, alert on, and understand their metrics no matter where they are stored. It also provides dashboards, plugins, alerts, reports, and integrations with various data sources and tools.
Influx database	An open source time series database that is designed to handle high write and query loads. It is used for storage and retrieval of time series data in fields such as operations monitoring, application metrics, Internet of Things sensor data, and real-time analytics.
MyCCI	The user portal of Competence Center ISOBUS e. V. (CCI) which is a registered association for promoting the technical development, accelerated take-up and international implementation of ISOBUS, a data transfer system for agricultural engineering.
Research and development	Research and Development (R&D) is the process of process and product innovation. It involves conducting research to discover new knowledge and applying it to create new or improved products, services, or processes. R&D is often done by businesses, universities, or government agencies to gain competitive advantage or solve problems.
Scrum	A framework for agile project management that focuses on delivering value in short iterations called sprints. Scrum involves a cross-functional team that works collaboratively to define and prioritize requirements, plan and execute tasks, review results, and adapt to changes. Scrum also uses roles, events, artifacts, and rules to guide the process.

Telemetry	The automatic recording and transmission of data from remote or inaccessible sources to a receiving equipment for monitoring and analysis. Telemetry is used in various fields such as meteorology, medicine, intelligence, and software development.
Trace files	Files that contain logs or records of the activity or performance of a system, program, or device. Trace files can be used to diagnose problems, debug errors, optimize performance, or monitor behavior. Trace files can have different formats and extensions depending on the source and purpose.

2 Introduction

The machines produced by Kuhn transmit telemetry data which can reveal how the products behave when they are used daily, and how they can be adjusted to better suit the customers' needs. Telemetry data can also confirm the ideas and designs that the company has developed, and show the actual benefits of making changes to the products. That's why the research and development department (R&D) in Geldrop uses this data in the process of designing new machines.

These machines are equipped with a GPS tracker and sensors that measure various parameters, such as the rotation speed and load of a plunger. These parameters are sent to a bus, a network that allows communication between different parts of the machine. The CAN logger tracks the data from the bus and stores it in a trace file. The CAN logger then sends the trace file via the mobile data network to a web portal called MyCCI.

Currently, the designers use the telemetry data that is being produced by test machines. To use the data that is actually being generated by the machines that are being used by the customers, the designers would have to manually process the data from each machine. This would result in scalability issues, as it is very time consuming to this for each machine.

The goal of this project is to develop a tool that retrieves telemetry data from the portal, converts it to a suitable format and then uploads it to the database. The data can then be viewed and used by the developers of the R&D team.

3 Background

Kuhn produces agricultural machinery with factories on 3 continents, distribution subsidiaries and importers covering over 100 countries. It was founded in 1828 by Joseph Kuhn near Saverne in France and currently employs over 6000 people. The primary activities include designing, manufacturing, and marketing a wide range of innovative, high-quality agricultural equipment and services to meet the diverse needs of agriculture worldwide. This includes machinery for crop cultivation, seeding and fertilizing, crop protection, hay and forage harvesting, livestock feeding, and landscape maintenance.

Due to the increasing world population, the demand for food is increasing. It is expected that by 2050 the world population will reach 9.8 billion people, which is an increase of approximately 60% compared to the world population in 2000. Kuhn's mission is to meet this rising demand for food by providing the agricultural sector with more efficient agricultural machines.

KUHN operates in the agricultural machinery sector holding over 2000 patents and achieving a yearly revenue of 1.5 billion euros, where it encounters competition from various companies. Notably, John Deere, AGCO, and CNH are among the competitors. These companies are considered 'full liners', meaning they manufacture both tractors and machines. In contrast, KUHN specializes exclusively in machine production. This specialization allows KUHN to focus entirely on designing and manufacturing high-quality, innovative agricultural machines.

In order to compete, it is important that Kuhn remains innovative. That is why Kuhn spends 4% annually on research and development. The production site in Geldrop features a research & development department. When the engineers from this department are designing new machines, telemetry data from test machines is being used. The ideal situation would be using telemetry data transmitted by machines that the customers are using, so that the machines that are being designed can be adjusted to suit the customer' needs, based on how the machines are used in daily life.

4 Project definition

4.1 Problem definition

Real-life telemetry data shows how the machines are used by the customers. This data helps the designers and engineers of the R&D department to adjust new machines in a way that suits the customers' needs. By using the data during the design process, the machines will be able to do what the customers need and prevent implementing unnecessary features which results in lower manufacturing costs.

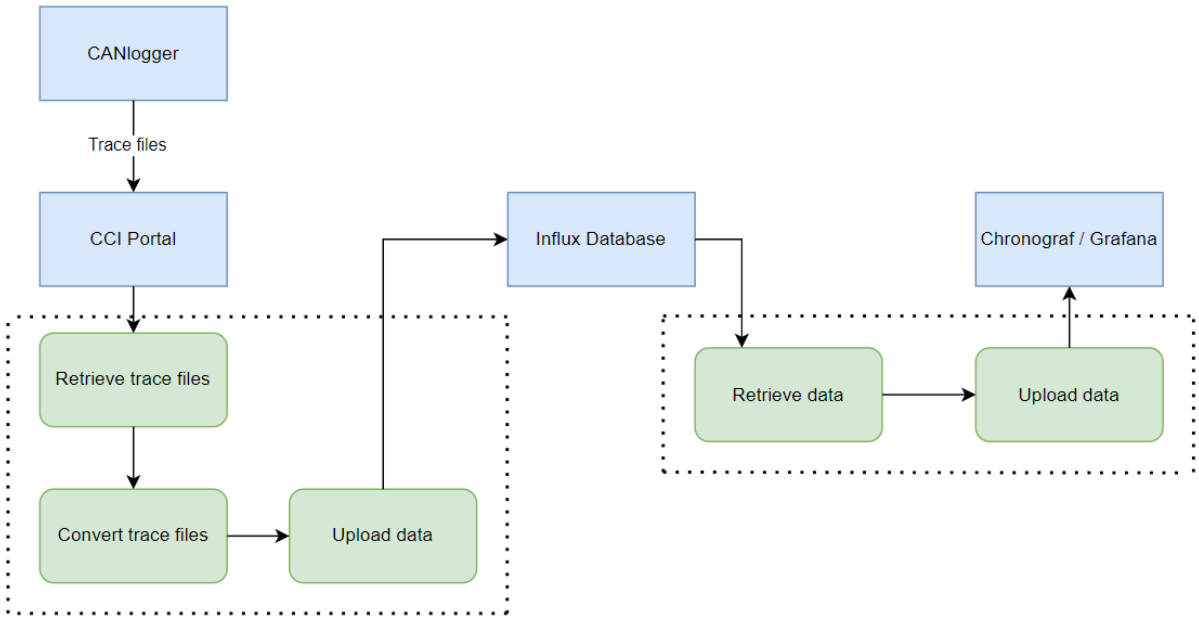

Figure 1 - Current situation

In Figure 1 the current process of how the telemetry data is being processed by the user is shown. The user first retrieves the file, that has been transmitted by the CANlogger inside the machine, from the MyCCI Portal. Next, the trace file is converted to an excel file. This is being done using a script that has been written by one of the software developers inside the R&D team. Then the Excel file can be uploaded to the Influx Database. To visualize the data, the user will have to download the data from the database and upload the data to Chronograf and/or Grafana.

The current situation has some bottlenecks that hinder the data collection and analysis process. There is no link between the MyCCI Portal and the Influx Database, so the user will have to manually convert each trace file. The existing scripts to convert the trace files are very specific, which means that if a change needs to be made to the format the user would have to edit the script. That would be very difficult if the user isn't familiar with programming. There also is no link between the database and Chronograf/Grafana.

Because of these scalability issues, the need for an automated data collection and analysis process is more and more prevalent in the R&D department.

4.2 Project goals

The objective of the project is as follows: “develop, within the internship period of 20 weeks, an application with which trace files can be processed and their data interpreted by non-programmers”.

4.3 Research questions

The main question of the project is as follows: “How can the data collection and analysis of the telemetry data become automated?”
To answer the main question, the following sub-questions have been formulated:

1. How can the MyCCI portal be linked with the Influx database?
2. How can the conversion of the trace files become automated?
3. How can the Influx database be linked with a front-end like Chronograf/Grafana?
4. How can the converted data be visualized and interpreted by a user in a non-programmer friendly way?

4.4 Project environment

To identify who is involved in automating the data collection and analysis, a stakeholder analysis was conducted. It looks at what interests each stakeholder has and how much influence they can have on the project. The result of the stakeholder analysis is to identify organizations that play a role in the project's decision-making or organizations that do not have a direct role in decision-making but can influence the project's progress. That influence can be either positive or negative.

Stakeholder	Primary/Secondary	Internal/External	Involvement and expectations	Approach to keep involved
Intern (Burak Arslan)	Primary	Internal	The intern performs the assignment and strives to produce a high-quality final product that is useful to the R&D department	-
Internship client (Stefan Polman)	Primary	Internal	The client provides guidance to the intern and expects a useful final product	At least a weekly meeting about the progress of the project
Internship supervisor	Primary	External	The internship supervisor monitors progress and expects commitment from the intern	A meeting at least once in two weeks, and multiple visits to the company throughout the internship
R&D Engineers	Secondary	Internal	The engineers of the R&D department are the users the final products and expect an automated data collection and analysis process that saves them time	Keep informed about the progress and make the products available when the project is finished
IT Department Saverne	Secondary	External	The colleagues in Saverne also experience the same issues and expect useful information	Make the products and documentation available when the project is finished

Table 1 – Stakeholders

In Table 1 the stakeholders of this project are shown, along with the involvement and expectation and the planned approach for each stakeholder.

4.5 Project approach

This paragraph describes the methods used to achieve the objective and intended result:

- The project will follow an agile methodology, which is based on the principles of iterative and incremental development, where requirements and solutions evolve through collaboration between self-organizing cross-functional teams. Agile methodology allows for flexibility and adaptability to changing requirements and feedback, as well as delivering value to the customer in short cycles.

- The project will use Scrum as a specific framework for implementing agile methodology. Scrum is a lightweight framework that helps teams work together and deliver products in complex environments. Scrum consists of three roles: the product owner, who represents the customer and defines the product vision and backlog; the Scrum master, who facilitates the Scrum process and removes impediments; and the development team, who are responsible for delivering potentially shippable increments of the product at the end of each sprint. The details of how the scrum framework will be applied is described in Communication plan.

- The project will be a solo project, meaning that there will be only one person who develops the application. The intern will work closely with a mentor or a coach, who will act as a Scrum master and a product owner, providing guidance, support, and feedback throughout the project. The mentor or coach will also help the intern plan and execute sprints, review the product increments, and conduct retrospectives.

- The project will use version control via GitHub, which is a web-based platform that hosts repositories of code and facilitates collaboration among developers. Version control allows the solo developer to track changes, manage branches, merge conflicts, and revert to previous versions of the code if needed. GitHub also provides features such as issue tracking, code review, pull requests, and GitHub Pages, which can help the solo developer communicate with the mentor or coach, improve the quality of the code, and showcase the product.

4.6 Scope & Exclusions

To get the boundaries of the project clear, the scope is defined. This means defining what will and will not be done during the project. Delineating the scope has several advantages. It clarifies the expectations of the client, the supervisor and the student. It ensures the feasibility and quality of the project. And it monitors the planning and progress of the project. Delineating the scope allows the student to focus on the main objectives and questions of the project without getting lost in irrelevant or unnecessary details.

What falls within the scope:

- Data from the trace files are converted to the correct format.
- The code written in a way that it is highly readable and maintainable by other programmers.
- Documenting and reporting the development process and results.
- Testing and evaluating both tools based on predefined criteria and indicators.

What falls outside the scope:

- The design of the graphical tool.
- Analyzing or interpreting the data from the Influx database. This is a task for the R&D department or other experts. The tools only provide the data in the desired format.
- Modifying or improving existing data collection tools, such as the CAN logger or web portal. This is the responsibility of the client or other parties.

4.7 Principles & Preconditions

Principles

The principles of this project are:

- The project will be developed in C# (.NET framework)
- The product has to be compatible with all the machine types and software versions that the trace files from MyCCI originate from
- The project will be completed in 20 weeks

Preconditions

The following preconditions apply to this project:

- The trace files have a standardized format and contain all the necessary information for the conversion
- The myCCI API is reliable and up-to-date
- The Influx database is accessible and secure
- The user has sufficient technical knowledge to use the CLI and the UI
- Chronograf and Grafana offer sufficient options to visualize the data

4.8 Available resources & dependencies

The project requires programs to help develop and test the code. Visual Studio and Visual Studio Code were chosen as the main development environments because they offer many features and extensions that make programming easier and more efficient. Visual Studio is an integrated development environment (IDE) that provides support for various programming languages, debugging, testing, versioning and collaboration. Visual Studio Code is a lightweight code editor that also supports many languages, as well as intellisense, syntax highlighting, code formatting, refactoring and live share.

In addition to these programs, DTC files are also required to determine the conversion configuration for trace files. DTC files are files that describe the structure and format of the data in the trace files. These files are essential to correctly read and analyze the trace files. There is dependence on the trace files coming from the machines because they contain information about the machines' performance, behavior and possible errors. By using these files, the code can be optimized and improved to meet the requirements of the project.

5 Project organizational structure

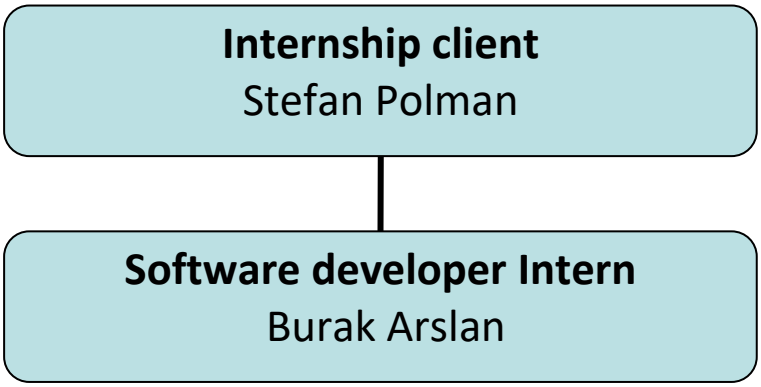

Table 2 - Project organizational chart

5.1 Internship client

Roles: Trainer, counselor, assessor

Tasks and responsibilities: Assessing the suitability, sustainability and relevance of the graduation assignment is one of the responsibilities of the internship client within the role of the supervisor. As a counselor the internship client must also advise and support the intern in determining, carrying out and estimating the assignment. Furthermore, the internship client must provide feedback and tips on the technical, theoretical and process aspects of the project. The intern must be inspired and challenged by the internship client to get the best out of himself. Additionally, the internship client must check and confirm whether the graduating student meets the requirements and criteria of the program and the graduation company. The achievements, results and conclusions of the intern will be assessed by the internship client.

5.2 Intern

Roles: Developer, researcher, advisor

Tasks and responsibilities: Developing a software product or software solution that meets the wishes and requirements of the client and the course is the main responsibility of the intern. The intern must also comply to the agreed schedule, quality and scope of the project. Furthermore, the intern must conduct research into the relevant technologies, frameworks, methods and techniques required to realize the assignment. The progress, results, conclusions and recommendations of the project must be reported and presented by the intern. Collaboration with the graduation supervisor, the client and any other parties involved is also expected from the intern. Additionally, the intern must reflect on his or her own learning process, competencies, performance and areas for improvement.

5.3 Organizational chart

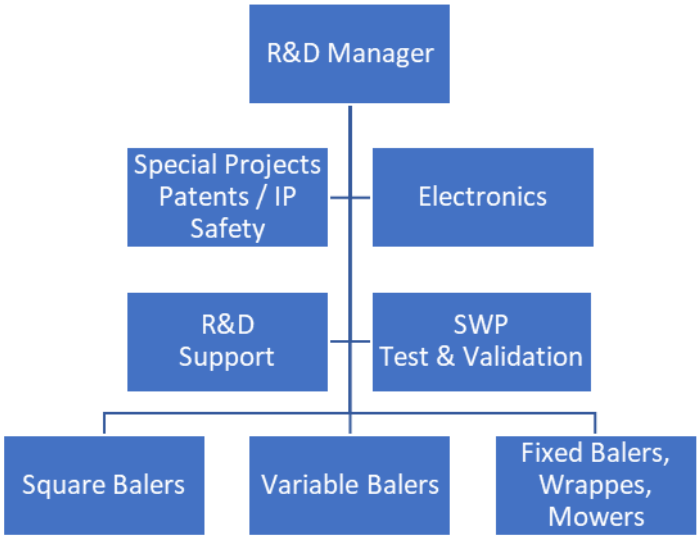

Table 3 - Organizational chart R&D department

In the table above, the organizational chart of the R&D department of the Kuhn division in Geldrop is shown. The departments that will be most affected by this project are electronics and SWP Test & Validation, since the designers will make decisions based on the data from machines that are in the field.

6 Products

6.1 Product breakdown structure

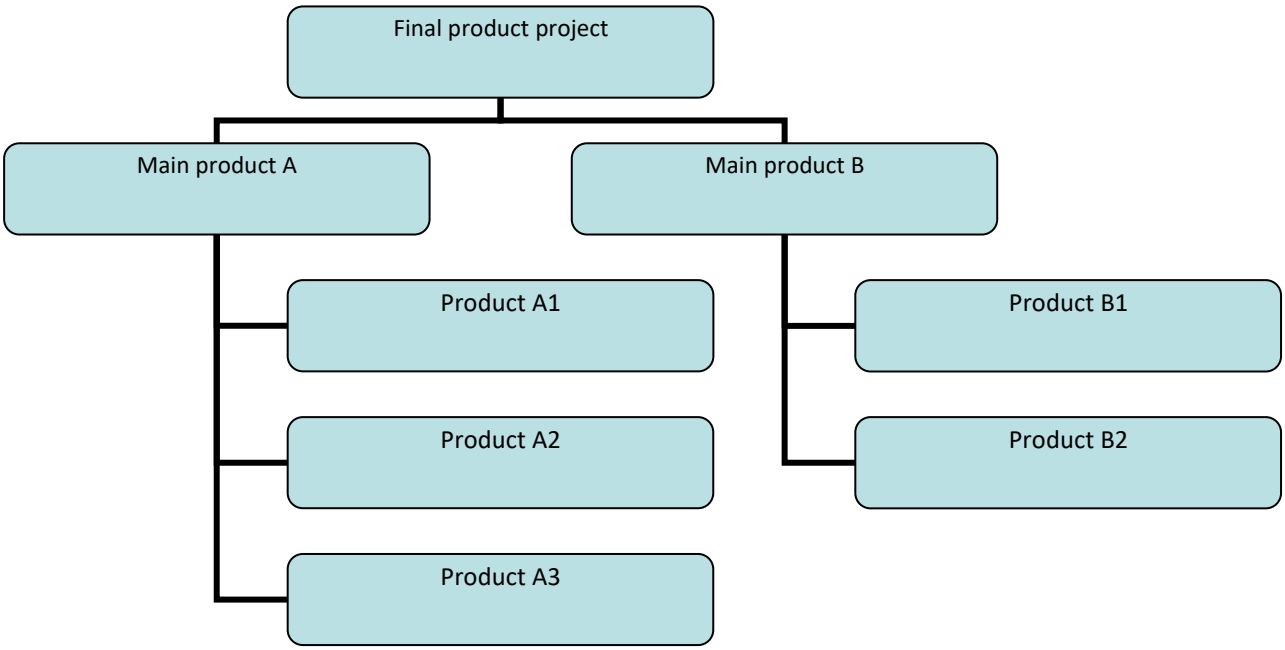

Table 4 - Product breakdown chart

The products that are going to be delivered are shown In Table 4 - Product breakdown chart. There will be two main products, which have been named Main product A and Main product B. Main product A consists of three subproducts and Main product B consists of two subproducts.

6.2 Product flowchart

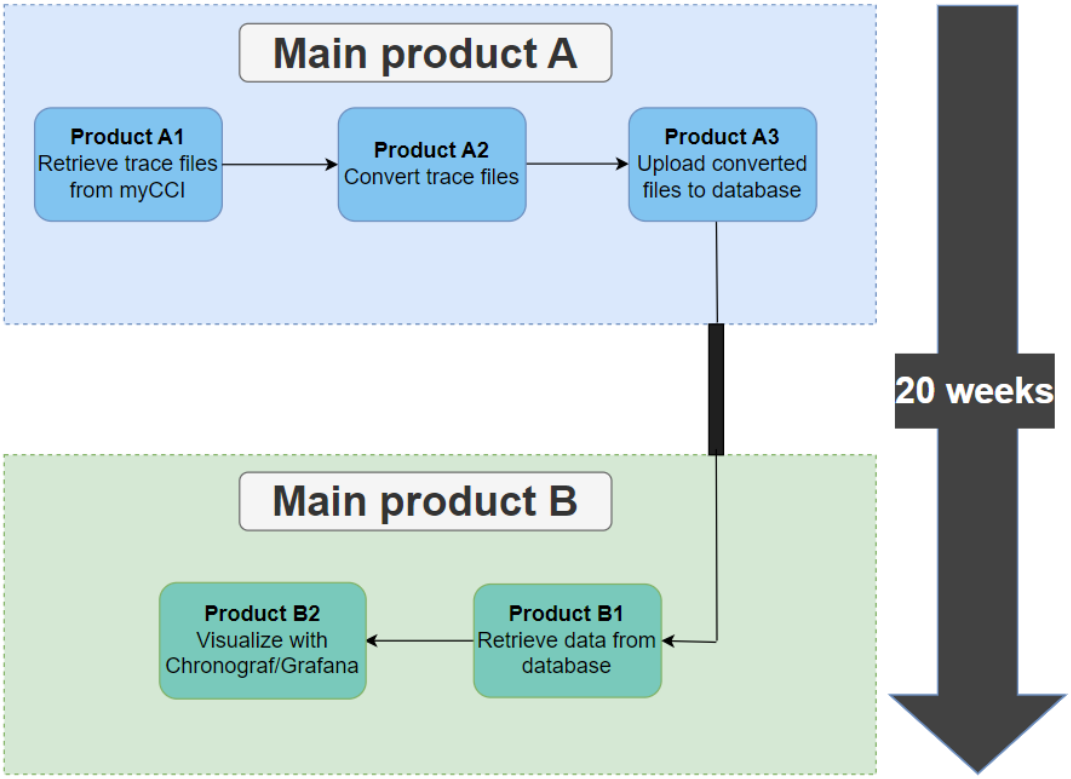

Figure 2 - Product flowchart

The product flowchart, as shown in Figure 2 - Product flowchart, has been made to show how the products will be made within the internship period of 20 weeks. The priority is on automating the processing of trace files, so Main product A will be developed first.

6.3 Product descriptions

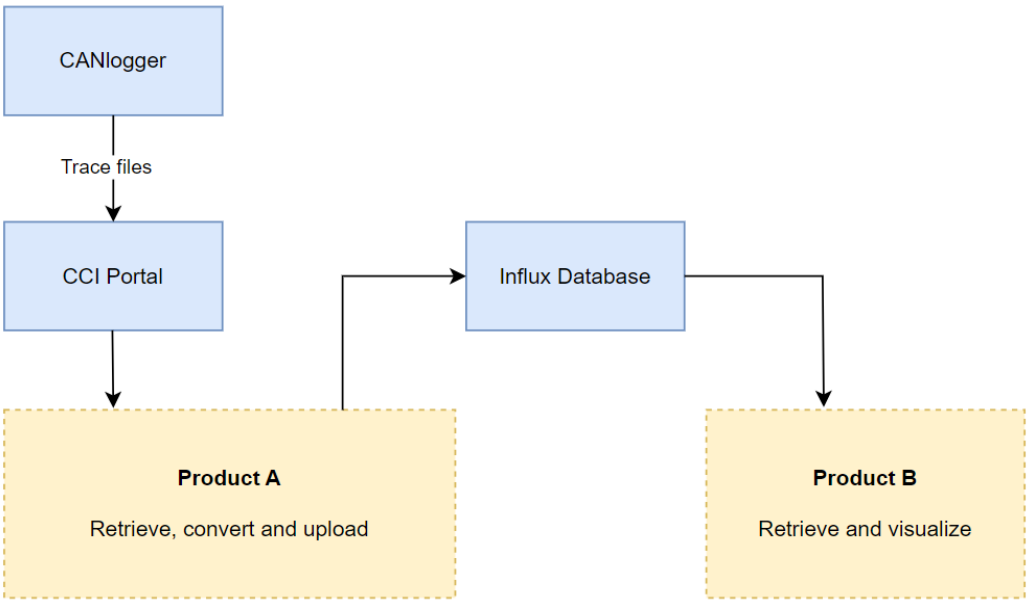

Figure 3 - Desired situation

After the products have been developed, the new situation will be like shown in Figure 3 - Desired situation. No user will have to access the MyCCI portal, since every trace file will be automatically retrieved, converted and uploaded to the Influx database. The user can apply certain filters In product B, so the user can analyze data that is relevant to the needs of the user.

The exact description and composition of the products are described in the tables below.

Main product A

	Main product A
Goal	Retrieve trace files from MyCCI, convert the files to a desired format and upload the files to the Influx database R&D employees will be able to process telemetry data on a large scale and won't have to do this manually anymore and will save time.
Composition	This product consists of a command line application developed with .NET (C#), which means that there is no UI.
Sources	
Quality criteria	When a machine transmits data to the MyCCI portal, in the form of a trace file, the product automatically retrieves the trace file - The product converts the trace file to the correct format, based on the machine type and software version - The product uploads the converted file to the Influx database
Approval	
Distribution	

	Product A1
Goal	When a trace file is transmitted to MyCCI, this product automaticcaly retrieves the file and stores is (temporarily) for convertion
Composition	Script in C# (.NET)
Sources	
Quality criteria	The product automatically retrieves trace files
Approval	
Distribution	

	Product A2
Goal	Convert the retrieved trace files to appropiate format.
Composition	Script in C# (.NET)
Sources	
Quality criteria	The product converts the trace file to the appropiate format and a JSON format. The JSON files allow the conversion to be checked
Approval	
Distribution	

	Product A3
Goal	Upload converted files to the database
Composition	Script in C# (.NET)
Sources	
Quality criteria	The product uploads the converted trace file to the Influx database
Approval	
Distribution	

Table 5 - Details of Product A

Main product B

	Main product B
Goal	Retrieve trace files from MyCCI, convert the files to a desired format and upload the files to the Influx database. R&D employees will be able to process telemetry data on a large scale and won't have to do this manually anymore and will save time.
Composition	This product consists of a front end in the form of a web app. The functionalities are based on Chronograf and Grafana.
Sources	
Quality criteria	User can apply filters to retrieve data for e.g. specific types of machines Based on the applied filters, the correct data is retrieved from the Influx database The user can generate various graphs/charts to interpret the data
Approval	
Distribution	

	Product B1
Goal	The product retrieves and visualizez the data, according to the wishes of the user
Composition	A front end based on the functionalities of Chronograf and Grafana. Created with React
Sources	
Quality criteria	The user can apply filters to get specific data The product retrieves the specific data by executing querys to the Influx database The product visualizes the data that has been retrieved
Approval	
Distribution	

Table 6 - Details of Product B

6.4 Job profile

The job profile is software developer. Most of the criteria of the function profile are met in the assignment. Here are some examples:

- A functional and maintainable application is designed and built with the most suitable technology. C# and .NET Framework are used for the graphical tools, because this is the standard technology at the client. Accountability can be given for the design and technology choices in the documentation and presentation.
- The importance of privacy and security aspects during software development is understood and these aspects are taken into account in the development activities. The trace files and the Influx database are only accessible for authorized users. Encryption and hashing are also used to secure the data.
- Insight is gained on how to perform maintenance within an existing software system driven by new functionalities or bug fixes. Version control and bug tracking tools are used to manage and improve the code. The coding standards and best practices of the client are also followed.
- Software quality is ensured by peer reviews and writing relevant tests at different levels. The code is regularly reviewed by the supervisor and other developers. Unit tests, integration tests and system tests are also written to test the functionality and the performance of the application.
- White box tests are designed, implemented and maintained. A test framework will be used to test the code on different scenarios and edge cases. Code coverage tools are also used to measure how much of the code is tested.
- Reviews are made of partial products such as: designs and code (including tests). Feedback is given on the designs and code of fellow students and colleagues. Feedback on own work is also welcomed and learned from.

7 Schedule

7.1 First 8 weeks

In the first period of the project, the main focus will be on acquiring the necessary knowledge to carry out the assignment. The following topics will be explored:

- How to automatically convert a trace file (hex data) to a readable format such as excel? Which format is the most suitable? The programming language C# and the DTC file that contains the meaning of the hex data will be used for this purpose. Different scripts will be written and tested to optimize the conversion.
- How to automatically upload the converted trace file to the database? How to also export the files to a JSON file, which can later be reviewed? How and where to store the JSON files?
- What are the possibilities of Influx Database to store and analyze data? The tools Chronograf and Grafana, which are used to visualize and edit data, will be familiarized with. The main question about this is how the UX can be suited for a non-programmer.
- What are the functionalities of MyCCI, the platform that is used to collect and send data? The documentation and manuals will be studied to understand how MyCCI works and how it can be used for the project. Contact could also be made with the developers of MyCCI to ask any questions or give feedback.
- To find out the needs of different users, interviews and surveys will be conducted. This will give a good idea of what the application's priorities should be in terms of functionalities.

The project plan will be written in the first four weeks. The project plan will include the objectives, scope, approach, deliverables, schedule of the project.

By acquiring this knowledge and writing the project plan, the activities in the next 12 weeks of the project will be well prepared.

7.2 Week 8 - 20

Product	Number of weeks	Week	1	2	3	4	5	6	7	8	9	10	11	12
A1	2													
A2	2													
A3	2													
A	**7**													
B1	2													
B2	2													
B	**5**													

Table 7 – Schedule week 8 - 20

In the following 12 weeks, the products will be developed. Product A will be developed first, followed by product B. There will be approximately 2 weeks needed for each partial product. After each partial product has been developed and working there will be 1 week to test and optimize the main product. As seen in the table above, product A will be developed in 7 weeks, and product B in 5 weeks.

8 Communication plan

8.1 Scrum framework

Kuhn's electronics team works with the scrum methodology. The intern will also participate with the sprints which means that the intern uses items, which describe the various tasks and activities needed to achieve the project goal. These items are placed in a backlog, where they are prioritized and estimated based on their value and complexity. For each sprint, a number of items are selected from the backlog to execute, taking into account available time and resources. When an item is finished, it is moved to the done column and another item can be picked up. In this way, the project is worked on iteratively and incrementally, with regular feedback and evaluation.

8.2 Sprints

As the project will be executed with scrum methodology, there will be a delivery on Wednesday every 2 weeks. So there will be a total of 6 deliveries in the 12-week period. The dates on which the delivery will take place can be seen in Table 8.

Delivery of sprint	Date
1	25-10-2023
2	8-11-2023
3	22-11-2023
4	6-12-2023
5	20-12-2023
6	10-1-2024

Table 8 - Delivery dates

8.3 Retrospective

To improve the quality and efficiency of the project, a retrospective will be held after each delivery. A retrospective is a time to look back on the work process and learn from the successes and mistakes. The following steps will be followed for each retrospective:

- An online tool will be used to choose a retrospective format, such as a starfish, a kite or a racing car. These formats will help to discuss different aspects of the project, such as what to start, stop, continue, do more or less.

- The supervisor, client or other stakeholders will be invited to participate in the retrospective. Their feedback on the work and their expectations for the next sprint will be asked for.

- The results of the retrospective will be summarized and documented. The key learning points and action points will be recorded and used to adjust and improve the project plan.

9 Bibliography

1. *About Kuhn*. (n.d.). Retrieved from kuhn.com: https://www.kuhn.com/en/about-kuhn
2. *Frequently Asked Questions*. (sd). Retrieved from MyCCI: https://mycci.cc-isobus.com/en/faq/
3. Rehkopf, M. (sd). *Scrum Sprints: Everything You Need to Know.* Retrieved from Atlassian: https://www.atlassian.com/agile/scrum/sprints
4. *World Population by Year*. (2023). Retrieved from worldometers.info: https://www.worldometers.info/world-population/world-population-by-year/
5. *World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100*. (2017). Retrieved from Website of the United Nations : https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100