**Vorverarbeitung und Aufbereitung von Dauerlauf-Messdaten aus LKW-Fahrzeugloggern zur Bereitstellung für interne Kunden bei Daimler Truck in Wörth**

Seminararbeit

**Maximilian de la Fuente Klein**

Matrikelnummer: 3583928

1.Betreuer: Martin Reißel

2.Betreuer: Andreas Weiland

**measX GmbH & Co. KG**

Dezember 20, 2024

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Unterschrift Auszubildender Unterschrift Betreuer

Inhalt

[1 Einleitung 3](#_Toc185589875)

[1.1 Ziel der Arbeit 3](#_Toc185589876)

[1.2 Projektkontext und Hintergrund 3](#_Toc185589877)

[1.3 Motivation und Relevanz für Daimler Truck 4](#_Toc185589878)

[2 Allgemeine Projektvorstellung 5](#_Toc185589879)

[2.1 Einführung in die Loggerdaten-Verarbeitung bei Daimler Truck 5](#_Toc185589880)

[2.2 Anforderungen und Spezifikationen 6](#_Toc185589881)

[2.3 Funktionale und technische Randbedingungen 7](#_Toc185589882)

[3 Datenaufbereitung und Vorverarbeitung 8](#_Toc185589883)

[3.1 Beschreibung der Loggerdaten und Formate 8](#_Toc185589884)

[3.2 Anforderungen an die Verarbeitung und Filterung 9](#_Toc185589885)

[3.3 Verzeichnisstruktur und Dateibenennung 10](#_Toc185589886)

[3.4 INI-Datei: Parametrierung und Konfiguration der Software 11](#_Toc185589887)

[4 Merger: Aufbau und Funktionsweise 11](#_Toc185589888)

[4.1 Überblick über den Merger-Prozess 11](#_Toc185589889)

[4.2 Architektur des Mergers 12](#_Toc185589890)

[4.3 Komponenten des Mergers 12](#_Toc185589891)

[4.3.1 Interface 13](#_Toc185589892)

[4.3.2 Poller 13](#_Toc185589893)

[4.3.3 Reader 13](#_Toc185589894)

[4.3.4 Writer 13](#_Toc185589895)

[4.3.5 JobManager 13](#_Toc185589896)

[4.4 Datenspeicherung und -sicherheit 13](#_Toc185589897)

[5 Prozessierung der Daten 14](#_Toc185589898)

[5.1 Scan der Importverzeichnisse und Erkennung von Datenformaten 14](#_Toc185589899)

[5.2 Zuordnung der Daten zu Tagesdateien 14](#_Toc185589900)

[5.3 Integration der Kanalnamen und Zeitkanäle 15](#_Toc185589901)

[6 Protokollierung und Fehlerbehandlung 15](#_Toc185589902)

[6.1 Log-Dateien und Fehlerprotokollierung 15](#_Toc185589903)

[6.2 Error.Log: Struktur und Inhalte 16](#_Toc185589904)

[6.3 Einfluss fehlerhafter Dateien auf den Gesamtprozess 17](#_Toc185589905)

[7 Auswertung (Statistik) 17](#_Toc185589906)

[7.1 Analyse und Auswertung der verarbeiteten Daten 17](#_Toc185589907)

[7.2 Statistische Verfahren zur Überprüfung der Datenqualität 18](#_Toc185589908)

[7.3 Grafische und tabellarische Darstellung der Ergebnisse 19](#_Toc185589909)

[8 Zusammenfassung und Ausblick 20](#_Toc185589910)

[9 Literaturverzeichnis 22](#_Toc185589911)

# Einleitung

## Ziel der Arbeit

Das Ziel dieser Seminararbeit ist die Beschreibung der Mitarbeit an der Entwicklung und Beurteilung eines effizienten Systems zur Vorverarbeitung und Aufbereitung von Messdaten aus Lkw-Fahrzeugloggern. Im Zusammenhang mit den Dauerlaufmessungen bei Daimler Truck in Wörth zielt diese Arbeit darauf ab, zu zeigen wie Messdaten für unterschiedliche interne Abteilungen automatisiert erstellt und zusammengefasst werden können. Durch die Software-Applikation, den „Merger“, erfolgt die Aufbereitung der Rohdaten in übersichtliche Tagesdateien in unterschiedlichen Dateiformaten pro Fahrzeug.

Das Hauptaugenmerk der Arbeit liegt auf der Entwicklung eines standardisierten Prozesses zur Datenvorverarbeitung. Es werden die Anforderungen und technischen Schwierigkeiten sowie die funktionale Umsetzung und Auswertung der Lösung beschrieben. Die Grundlage wird gegeben, durch eine genauere Betrachtung der eingesetzten Datenformate, Schnittstellen und der Verarbeitungsschritte, die für eine einheitlichen Datenqualität und strukturierten Speicherung gefordert sind.

Zudem werden die Beschränkungen und erforderlichen Annahmen berücksichtigt, die im Laufe des Projekts festgelegt wurden. Dabei ist es besonders schwierig, die Datenverarbeitung zu vereinfachen, da spezielle Fälle und aufwendige Plausibilitätsprüfungen bewusst nicht implementiert werden sollen. Daher soll eine Basisapplikation bereitgestellt werden, die die wesentlichen Anforderungen an die Datenaufbereitung erfüllt und als Grundlage für künftige Erweiterungen des Projekts dient.

## Projektkontext und Hintergrund

Daimler Truck in Wörth sammelt Dauerlaufmessungen, die kontinuierlich umfangreiche Daten aus verschiedenen Lkw-Fahrzeugloggern sind. Diese Langzeitmessungen helfen bei der Erstellung von Erkenntnissen über die Fahrzeugleistung und –Belastbarkeit im tatsächlichen Betriebszustand. Interne Kunden müssen auf die verschiedenen Messparameter zugreifen können, um technische Analysen und Optimierungen auf Grundlage die erfassten Loggerdaten durchzuführen.

Es existieren zwei wichtige Datenformate für die gesammelten Messdaten: konventionelle Loggerdaten und SOME\_IP-Daten, sowie separate und kombinierte Auswertungen. Jeder Fahrzeugtyp hat sein eigenes Importverzeichnis für die Daten, die mit den Daten aus dem entsprechenden Fahrzeug leicht verknüpft werden können. Dies bezieht sich auf die Rohdateien, die anhand eines von Daimler vorgegebenen Musters, jedoch auf die Spezifität der Zeitstempel im Dateinamen, bereitgestellt werden müssen. Daimler liefert die Daten im MDF4-Format. Jede Datei beinhaltet mehrere Tausend Kanäle, die alle unterschiedlichen Messungen darstellen.

Die Loggerdaten werden über den CAN-Bus (Controller Area Network) erfasst, der speziell für die robuste und zeitkritische Datenübertragung im Fahrzeug entwickelt wurde.

Die SOME\_IP-Daten basieren auf dem Ethernet-Bus, der für die Übertragung großer, datenintensiver Datenmengen ausgelegt ist.

Eine manuelle Verarbeitung dieser Mengen würde zu lange dauern und würde nach wie vor zu vielen Ungenauigkeiten führen. Da pro Zeitabschnitt, zwischen Starten und Stoppen der Zündung ein Datenpaket geliefert wird, müssen meistens viele Datenpakete zu einer Tagesdatei zusammengefasst werden. Aus diesem Grund ist es von zentraler Bedeutung, die Vorverarbeitung zu automatisieren. Die Schwierigkeit besteht dabei aber, eine Software zu entwickeln, die die Messdaten aller Fahrzeuge auf einmal, ordentlich korrekt zuordnet und nicht viele aufwändige Plausibilitäts- oder Fehlerkorrekturvorgänge fordert. Bei Datenfehlern oder abweichenden Formaten wird eine manuelle Anpassung durch technisches Personal erwartet, was die Anwendung auf die standardisierten Anforderungen und definierten Datenformate fokussiert.

Für Daimler Truck ist es entscheidend, dass die Basisapplikation – auch „Merger“ genannt – die Rohdaten in chronologisch aufeinanderfolgende Tagesdateien je Fahrzeug konsolidiert und zusammenfasst. Diese Dateien werden anschließend aufbereitet und internen Abteilungen für Analysezwecke zur Verfügung gestellt. Das Projekt soll sich nur auf die Kernanforderungen der Datenzusammenführung ausrichten, da aus Kostengründen erweiterte Funktionen wie Plausibilitäts- oder Konsistenzprüfungen und eine Benutzeroberfläche hier nicht berücksichtigt werden.

## Motivation und Relevanz für Daimler Truck

Die zuverlässige und übersichtliche Erfassung und Aufbereitung der Fahrzeugmessdaten sind für Daimler Truck von strategischer Bedeutung. Die Langzeitmessungen liefern entscheidende Einblicke in die wirklichen Bedingungen des Einsatzes und ermöglichen es, die präzise Analyse der Komponenten der Fahrzeuge und die Qualitätssicherung unserer Produkte zu gewährleisten. Die zielgerichtete Entwicklung neuer Innovationen wird ebenfalls unterstützt.

Dabei zielt das Projekt insbesondere darauf, die Komplexität und Bearbeitungszeit zu reduzieren. Die Automation der Datenverarbeitung führt zur Erzeugung der Tagesdateien durch den Merger, von denen alle internen Kunden profitieren. Außerdem ist das Ziel, eine gemeinsame Zeitbasis für alle Messdaten zu schaffen, da dies nicht bei jedem Datenkanal der Fall ist. Auch wenn hier die Veränderungen beschrieben sind, ist zunächst festzustellen, dass die Mitarbeiter\*innen aufgrund der notwendigen manuellen Interventionen entlastet werden und die Gefahr der Fehler gemindert wird. Auch die Agilität in der Analyse und Optimierung der Fahrzeugkomponenten wird dadurch signifikant gefördert.

In Zukunft bildet das Projekt die Grundlage für weitere Entwicklungen. Plausibilitätschecks, eine benutzerfreundliche Oberfläche sind nur zwei Beispiele dafür. Insgesamt bietet der Merger einen pragmatischen Ansatz für die bestehende Datenverarbeitung bei Daimler Truck und schafft gleichzeitig Erweiterungsmöglichkeiten, die die gestiegenen Anforderungen durch die steigende Anzahl von Transporten aufgreifen können.

# Allgemeine Projektvorstellung

Die Inhalte dieses Kapitels basieren auf den internen Unterlagen von *Daimler Wörth (2024).*

## Einführung in die Loggerdaten-Verarbeitung bei Daimler Truck

Die Erfassung und Verarbeitung von Logger-Daten spielt im Fahrzeugtestumfeld von Daimler Truck eine wichtige Rolle. Logger-Daten werden von sogenannten „Loggern“ erstellt, die während Langzeittests kontinuierlich Betriebsparameter. Diese Messwerte geben wichtige Informationen in Bezug auf das Verhalten und die Grenzen der Komponenten von Lkw im täglichen Betrieb, die zur Verbesserung und Optimierung der Fahrzeuge führen Entwicklung und Produktion. Durch systematische Auswertung der Logger-Daten unterstützen die Entwicklung und Qualitätssicherung darauf ausgelegte Prozesse – wie die Logger-Daten Verbesserungen und Pflege der Leistung, Sicherheit und Zuverlässigkeit der Fahrzeuge.

Logger-Daten bestehen im Allgemeinen aus einer Vielzahl von Messkanälen, einschließlich: Motor-, Bremse-, Fahrwerk-, Kommunikationsdaten zwischen den Komponenten des jeweiligen Fahrzeugs. Der kontinuierliche Echtzeit-Lesevorgang führt zu enormen und informativen Datenmengen in definierten Dateiformaten – meist MDF4. Die Dateien enthalten neben den Messwerten auch viele unterschiedliche Zeitkanäle, die eine chronologische Analyse und Weiterverarbeitung der Fahrzeugaktivitäten ermöglicht.

Um diese große Menge an intern für Analysen nutzbaren Daten sinnvoll einzusetzen, werden Parser erstellt, die den gleichen Prozess beherrschen und somit die Informationen strukturieren können. Für Daimler Truck stellt die Aufgabenstellung hohe und komplexe Erwartungshaltungen dar: Interne Datenintelligenz soll in einem einzigen, übergeordneten File durchsichtig und transparent gemacht werden, um die Ergebnisse bei Bedarf flexibel in darauf basierende Analyseroutinen zu integrieren. Dies bildet die Grundlage für tiefgehende Analysen, wie die Erkennung von Verschleißmustern oder die Prüfung neuer Fahrzeugtechnologien.

Die Merger-Applikation unterstützt diese Prozesse erheblich, indem sie die Daten verarbeitet und den internen Kunden zur Verfügung stellt. Die Zusammenführung der Loggdateien in Tages-Dateien bietet eine klare Trennung, die Informationen übersichtlich und schnell sicherstellt, dass die Dateien zu jeder Zeit bedarfsgerecht abrufbar sind, da eine schnelle Verfügbarkeit der aktuellen Daten für Daimler Truck von großer Bedeutung ist.

## Anforderungen und Spezifikationen

Die Inhalte dieses Kapitels basieren auf den internen Unterlagen von *measX GmbH & Co. KG (2024).*

Die automatisierte Verarbeitung von Loggerdaten bei Daimler Truck muss verschiedenen technischen und funktionalen Anforderungen gerecht werden, um eine zuverlässige und effiziente Bereitstellung für interne Analysezwecke zu gewährleisten. Der Schwerpunkt liegt dabei auf der Konsolidierung und Aufbereitung der Messdaten in chronologisch geordneten Tagesdateien pro Fahrzeug. Die Anforderungen an die Merger-Applikation sind speziell auf die standardisierten Bedürfnisse der verschiedenen internen Abteilungen und die technischen Rahmenbedingungen der Datenerfassung zugeschnitten.

Zu den funktionalen Anforderungen gehört die Trennung der zwei Datenformate, konventionelle Loggerdaten und SOME\_IP-Daten, die separat verarbeitet und in getrennten Ergebnisdateien pro Fahrzeug abgelegt werden. Jede Datei wird anhand des ersten Zeitstempels im Dateinamen einem spezifischen Kalendertag zugeordnet. Die Zusammenführung muss dabei sicherstellen, dass alle Dateien korrekt in den jeweiligen Tagesdatensatz integriert werden, ohne dass es zu zeitlichen Überschneidungen oder doppelter Datenablage kommt.

Technische Anforderungen:

* Verarbeitung und Ausgabeformate: Die Merger-Applikation muss Loggerdaten und SOME\_IP-Daten im MDF4-Format verarbeiten und eine Auswahl an Ausgabeformaten unterstützen: MDF4, TDMS oder MATLAB. Da die Kunden, die mit den Daten weiterarbeiten, in verschiedenen Bereichen und somit auch mit verschiedenen Tools arbeiten, müssen auch verschiedene Ausgabeformate gegeben sein. Diese Formate werden je Fahrzeug in einer INI-Konfigurationsdatei definiert, die es ermöglicht, die Verarbeitungsparameter und das gewünschte Ausgabeformat flexibel anzupassen.
* Datenvalidierung und Fehlerbehandlung: Bei fehlerhaften oder beschädigten Dateien stoppt die Applikation die Verarbeitung für das betroffene Fahrzeug und dokumentiert den Fehler in einer speziellen Log-Datei.
* Zugriffskontrolle und Datensicherheit: Um eine kontinuierliche Verarbeitung ohne Einschränkungen des Zugriffs zu gewährleisten, sollten die Tagesdateien nicht schreibgeschützt sein. Zur Sicherheit bekommt unser System Kopien der fertigen Daten.

Daimler pflegt eine zentrale INI-Konfigurationsdatei, die diese technischen Anforderungen steuert. Diese enthält spezifische Parameter für jedes Fahrzeug, darunter das Importverzeichnis, das Ausgabeformat und die Kanallisten für die Ergebnisdateien. Auf diese Weise ist es möglich, Anpassungen der Verzeichnisstrukturen oder der Ausgabeformatanforderungen flexibel vorzunehmen, ohne dass die Software selbst angepasst werden muss.

Zusammenfassend stellt die Merger-Applikation mit ihrer zuverlässigen Datenstruktur, den flexiblen Konfigurationsmöglichkeiten und der robusten Dateiverarbeitung eine wertvolle Unterstützung bei der effizienten Nutzung der Loggerdaten dar und ermöglicht eine flexible Bereitstellung der verarbeiteten Datenformate.

## Funktionale und technische Randbedingungen

Die Merger-Applikation besitzt klar definierten Rahmenbedingungen, die sich aus Kostengründen auf die Effizienz und Einfachheit fokussieren.

Funktionale Randbedingungen:

* Fokus auf den Standardfall: Die Daten sind korrekt und in chronologischer Reihenfolge
* Trennung der Datenformate: Loggerdaten und SOME\_IP-Daten werden getrennt verarbeitet.

Technische Randbedingungen:

* Datenformate: Nur MDF4-Dateien werden unterstützt. Andere Formate oder Dateinamen werden gelöscht.
* Parametrierung:Über eine INI-Datei werden Verzeichnisse, Ausgabeformate (MDF4, TDMS, MATLAB) und die Kanallisten angegeben.

Diese Rahmenbedingungen bilden die Grundlage für die effiziente und anpassbare Verarbeitung der Fahrzeug-Messdaten je nach Gebrauch.

# Datenaufbereitung und Vorverarbeitung

## Beschreibung der Loggerdaten und Formate

Die Messdaten, die bei den Fahrzeugtests bei Daimler Truck aufgezeichnet werden, liegen hauptsächlich in zwei Formaten vor:

* Loggerdaten: Diese sind Messdaten, die kontinuierlich während der Testläufe der Fahrzeuge aufgezeichnet werden. Sie enthalten viele verschiedene Signale, darunter Fahrzeuggeschwindigkeit, Motordaten, Umwelteinflüsse und weitere Parameter. Zudem beinhalten die Loggerdaten die Kommunikation zwischen Steuergeräten, die über den CAN-Bus (Controller Area Network) erfolgt. Der CAN-Bus ist speziell für die robuste und zeitkritische Datenübertragung innerhalb des Fahrzeugs ausgelegt und wird für grundlegende Fahrzeugfunktionen genutzt. Die Daten werden immer im MDF4-Format gespeichert. Das MDF4-Format (Measurement Data File Version 4) ist ein standardisiertes Dateiformat, das hauptsächlich in der Automobilindustrie verwendet wird, um Messdaten zu speichern. Es wurde von ASAM (Association for Standardisation of Automation and Measuring Systems) entwickelt und ermöglicht eine effiziente Speicherung großer Datenmengen. Die Loggerdaten werden in einer großen Gruppe gespeichert und alle auf denselben Zeitkanal abgebildet.
* SOME\_IP-Daten: SOME\_IP (Scalable Service-Oriented Middleware over IP) ist eine Kommunikationsarchitektur, die zur Übertragung großer, datenintensiver Informationen in einem Fahrzeug über den Ethernet-Bus genutzt wird. Diese Daten beinhalten hauptsächlich Statusmeldungen, Diagnosedaten und komplexe Informationen. Diese Daten werden ebenfalls im MDF4-Format gespeichert, jedoch durch den Zusatz "\_SOME\_IP" im Dateinamen von den normalen Loggerdaten unterschieden. Das MDF4-Format ermöglicht eine strukturierte Speicherung dieser Daten, sodass sie effizient erfasst, übertragen und weiterverarbeitet werden können, um die Leistung und Funktionalität des Fahrzeugs zu überwachen und zu optimieren.

Merkmale des MDF4-Formats:

* Effiziente Speicherung: Das MDF4-Format ermöglicht die strukturierte Speicherung großer Datenmengen.
* Flexibilität: Es unterstützt eine Vielzahl von Datentypen und -strukturen, was es perfekt für die Speicherung der Fahrzeugdaten macht.
* Kompatibilität: Das Format ist ein Standard in der Automobilindustrie und wird von vielen Analyse- und Simulationswerkzeugen unterstützt.

Diese klar strukturierte Benennung erleichtert die Zuordnung der Daten zu den bestimmten Fahrzeugen und Zeiträumen und dient als Grundlage für die automatisierte Verarbeitung durch den Merger.

## Anforderungen an die Verarbeitung und Filterung

Die Verarbeitung der Loggerdaten erfolgt unter festgelegten Anforderungen, die sicherstellen, dass die Rohdaten effizient aufbereitet werden:

* Trennung der Datenformate: Loggerdaten und SOME\_IP-Daten werden unabhängig voneinander verarbeitet und in separaten Tagesdateien gespeichert.
* Chronologische Zuordnung: Jede Datei wird anhand des ersten Zeitstempels im Dateinamen einem Kalendertag zugeordnet.
* Reduktion auf relevante Daten: Die Verarbeitung beschränkt sich auf die in der Kanalliste definierten Kanäle. Nicht definierte oder irrelevante Daten werden ignoriert.
* Fehlerhafte Dateien: Dateien mit ungültigen Formaten oder abweichenden Strukturen werden nicht verarbeitet und automatisch gelöscht. Man kann in einer Log-Datei einsehen, welche Dateien behandelt und welche gelöscht wurden.
* **Gemeinsame Zeitbasis**: Die Kanäle sollen alle auf einen gleichen, einheitlichen Zeitkanal abgebildet werden.
* **Einheitliche Abtastrate**: Alle Kanäle sollen eine einheitliche Abtastrate von 10hz haben, um die Auswertung einfacher zu gestalten.

Diese Anforderungen dienen dazu, die Komplexität der Datenverarbeitung zu reduzieren und den Fokus auf das relevante, also die Erstellung von Tagesdateien zu legen.

## Verzeichnisstruktur und Dateibenennung

Die Organisation der erhaltenen MDF4-Daten und Ergebnisdateien erfolgt in einer klar definierten Verzeichnisstruktur, die eine einfache Nachvollziehbarkeit und Handhabung ermöglicht:

Eingangsverzeichnisse:

* Für jedes Fahrzeug gibt es eigene Verzeichnisse, in denen die Loggerdaten und SOME\_IP-Daten enthalten sind.
* Die Daten werden anhand des Dateinamens automatisch erkannt, da sie immer die gleiche Struktur aufweisen:
	+ Loggerdaten: <Fahrzeug-ID>\_<Startzeit>\_\_<Endzeit>.MF4
	+ SOME\_IP-Daten: <Fahrzeug-ID>\_SOME\_IP\_<Startzeit>\_\_<Endzeit>.MF4

Ergebnisverzeichnisse:

* Ergebnisdateien werden nach Fahrzeug-ID und Datum organisiert. Es gibt für jeden Tag eigene Tagesdateien für Logger- und SOME\_IP-Daten.
* Beispiel für eine Ergebnisdatei: 123A456\_2024-10-30.MF4

Dateibenennung und Filterung:

* Dateien mit falschem Format oder abweichenden Namen werden automatisch gelöscht, um eine fehlerfreie Verarbeitung zu gewährleisten.
* Der erste Zeitstempel im Dateinamen, also der Anfang der Aufzeichnung, bestimmt den zugeordneten Kalendertag und wird dementsprechend einsortiert.

Diese Verzeichnisstruktur gewährleistet eine klare Trennung der Daten nach Fahrzeug und Aufzeichnungstag. Sie sorgt für eine effiziente Verarbeitung und eine einfache Nachverfolgbarkeit der Roh- und Ergebnisdaten.

## INI-Datei: Parametrierung und Konfiguration der Software

Die INI-Datei spielt eine große Rolle bei der Konfiguration der Merger-Applikation. Sie dient als Schnittstelle zwischen den einzelnen Sektionen. In der Datei werden alle notwendigen Parameter hinterlegt, um eine flexible und anpassbare Verarbeitung zu ermöglichen.

Struktur der INI-Datei:
Die INI-Datei ist in verschiedene Sektionen unterteilt, wobei jede Sektion die Einstellungen für ein spezifisches Fahrzeug definiert.

* Beispiel für eine Fahrzeug-Sektion:

[Fahrzeug\_ID]

Enable = True

InputFolder = Pfad/zum/Importverzeichnis

OutputFolder = Pfad/zum/Ergebnisverzeichnis

OutputFormat = MDF4, TDMS, Matlab

chnListFileName = Kanalliste.xlsx

Bei „Enable“ wird festgelegt, ob das Fahrzeug verarbeitet werden soll. Im „InputFolder“ liegen die zu bearbeitenden MDF4 Dateien und im „OutputFolder“ landen am Ende die zusammengefügten Ergebnisdateien. Bei dem „OutputFormat“ können sowohl einzelne Formate, also auch mehrere Gleichzeitig angegeben werden. Dann werden auch direkt die Dateien in mehreren Formaten zusammengefügt. Unter „chnListFileName“ wird dann schließlich noch der Name der Excel-Liste angegeben, wo die Kanalliste enthalten ist. Der Pfad dazu wird vorher schon festgelegt.

# Merger: Aufbau und Funktionsweise

## Überblick über den Merger-Prozess

Der Merger-Prozess fasst Logger- und SOME\_IP-Daten pro Fahrzeug und Tagesabhängig in eine Tagesdatei zusammen. Jede Datei wird basierend auf ihrem Zeitstempel dem entsprechenden Kalendertag zugeordnet.

Wichtige Schritte im Prozess:

* Datenidentifikation: Die Erkennung und Zuordnung von Dateien anhand des vorgegebenen Namensschemas.
* Zusammenführung: Die Daten aus mehreren Dateien eines Tages werden an eine Tagesdatei angehängt oder diese neu erstellt.
* Ergebnisformate: Die Speicherung der aufbereiteten Daten in den Formaten MDF4, TDMS oder MATLAB.
* Gemeinsame Zeitbasis: Pro Ergebnisdatei gibt es einen Zeitkanal, auf den alle Kanäle abgebildet sind.
* Einheitliche Abtastrate: Alle Kanäle haben die gleiche Abtastrate, indem fehlende Daten, passend zu den vorhandenen Daten, hinzugefügt werden.

Fehlerhafte Dateien werden gelöscht, protokolliert, jedoch wird die Verarbeitung des Fahrzeugs nicht gestoppt. Der Prozess ist automatisiert und ermöglicht eine schnelle Verarbeitung von großen Datenmengen. Sobald neue Daten im Importverzeichnis erscheinen, werden sie automatisch ihrer Tagesdatei zugeordnet und verarbeitet.

## Architektur des Mergers

Der Mergervorgang verfügt über ein modulares Design, das eine klare Trennung der Komponenten und leichtes Erweitern ermöglicht. Jede Komponente erfüllt eine spezifische Aufgabe, sodass die Daten effizient verarbeitet werden. Folgende Hauptteile des Mergers sind vorhanden: das Interface, der Poller, JobManager, verschiedene Reader je nach Format des einzulesenden Files und der entsprechende Writer. Die modulare Struktur erleichtert die genaue Fehlerbehandlung, das Hinzufügen neuer Funktionen und das Aktualisieren des Mergers im Allgemeinen.

## Komponenten des Mergers

Der Merger besteht aus mehreren zentralen Komponenten, die in einem strukturierten Zusammenspiel die automatisierte Verarbeitung der Logger- und SOME\_IP-Daten ermöglichen.

### Interface

Das Interface definiert eine RawChannelGroup, die die Grundlage bietet um die Daten zwischenzuspeichern. Sie beinhaltet die Properties, den Gruppennamen, den Zeitkanal und die einzelnen Kanäle mit ihren Daten.

### Poller

Der Poller überwacht kontinuierlich die definierten Importverzeichnisse auf neue Dateien und bemerkt, sobald eine neue Datei erkannt wird.

### Reader

Es gibt einzelne Reader für die Formate MDF4, TDMS und MATLab. Sie lesen aus den Rohdateien und auch aus den Mergedateien die einzelnen Daten, die in der RawChannelGroup gebraucht werden, um alles für das Schreiben vorzubereiten. Durch die passenden Python Bibliotheken funktioniert das recht schnell, auch bei großen Datenmengen. Bei MDF4 wird die Bibliothek asammdf verwendet. Bei TDMS ist es nptdms und bei MATLab scipy.io.

### Writer

Der Writer ist für die Erstellung und Aktualisierung der Tagesdateien verantwortlich. Er fügt neue Daten an bestehende Tagesdateien an oder erstellt neue Dateien, falls diese noch nicht existieren. Die Ausgabe erfolgt in den definierten Formaten (MDF4, TDMS, MATLAB).

### JobManager

Über den JobManager wird der ganze Mergeprozess gestartet und sorgt als Zusammenspiel der ganzen anderen Prozesse.

Diese Komponenten arbeiten alle zusammen, um die Datenverarbeitung effizient zu gestalten.

## Datenspeicherung und -sicherheit

Die letzten verarbeiteten Daten werden gemäß einer klaren Dateiverschlüsselung auf den Speichern abgelegt, um eine einfache Verwaltung und schnellen Zugriff zu gewährleisten. Jede Ergebnisdatei wird anhand der Fahrzeug-ID und des Datums identifiziert.

In der Zwischenzeit sollte bereitgestellt werden, dass die Ergebnisdateien jederzeit überlagert werden können, um die Aktualisierung zu ermöglichen. Dadurch wird die Datei schreibgeschützt, und das bedeutet, dass während des Verarbeitungsprozesses nicht geändert werden kann.

# Prozessierung der Daten

## Scan der Importverzeichnisse und Erkennung von Datenformaten

Die Merger-Applikation beginnt den Verarbeitungsprozess mit einem Scan aller Import- und deren Unterverzeichnisse. Für jedes Fahrzeug wird ein eigenes Verzeichnis überwacht, welches die Rohdaten enthält.

Der Poller identifiziert die zu bearbeitenden Dateien anhand ihres Namens und der Namensstruktur, um sicherzustellen, dass nur unterstützte Formate verarbeitet werden. Dabei gelten folgende Regeln:

* Unterstützte Formate: Nur Dateien im MDF4-Format werden erkannt und für die Verarbeitung zugelassen.
* Dateibenennung: Die Dateien müssen der vorgegebenen Struktur entsprechen, die Fahrzeug-ID und Zeitstempel enthält.
* SOME\_IP-Erkennung: Dateien mit „SOME\_IP“ im Namen werden separat als spezielle Daten gekennzeichnet.

Alle Dateien werden gelöscht, die nicht die erforderlichen Informationen enthalten, damit die Datenverarbeitung konform verläuft. Die gelöschten Dateien werden jedoch in eine Logdatei aufgenommen, mit der Sie einsehen können, welche Dateien gelöscht wurden.

## Zuordnung der Daten zu Tagesdateien

Nach der Erkennung von den neuen Dateien erfolgt die Zuordnung der Messdaten zu den entsprechenden Kalendertagen. Die Zuordnung erfolgt nach dem ersten Zeitstempel im Dateinamen, also dem Startdatum.

Beispiel:
Eine Datei mit dem Namen
123a1234\_2024-04-30\_07-36-32\_\_2024-04-30\_08-38-06.MF4
wird dem Kalendertag 30. April 2024 zugeordnet, auch wenn sie über mehrere Tage hinweg Daten beinhaltet.

Die Daten aus den Dateien eines Tages werden in einer gemeinsamen Tagesdatei zusammengeführt. Falls für den entsprechenden Tag noch keine Tagesdatei existiert, wird diese neu erstellt. Loggerdaten und SOME\_IP-Daten werden jeweils in separaten Tagesdateien verarbeitet.

## Integration der Kanalnamen und Zeitkanäle

Während der Verarbeitung werden die Messkanäle aus den Rohdaten anhand der Vorgaben der INI-Datei in die Tagesdateien integriert. Für jeden Datensatz wird ein einziger Zeitkanal hinzugefügt, der für alle weiteren Kanäle zutrifft.

* Zeitkanal: Ein absoluter Zeitkanal mit einer Auflösung von 10 Hz wird für jede Tagesdatei generiert. Er stellt sicher, dass die Daten chronologisch korrekt dargestellt werden und keine Lücken innerhalb der Kanäle entstehen, falls einzelne Kanäle mit einem anderen Zeitkanal getaktet wurden. Wenn ein Kanal anders getaktet ist, werden die Lücken einfach aufgefüllt, um bei der Auswertung keine Sprünge innerhalb der Daten zu haben.
* Kanalnamen: Nur die in der Kanalliste enthaltenden Kanäle werden in die Ergebnisdatei geschrieben. Diese Liste wird für jedes Fahrzeug individuell definiert.

Die Messdaten aus den Rohdaten werden auf den Zeitkanal abgebildet, was sowohl die Datenanalyse als auch den Zugriff auf einzelne Kanäle erleichtert.

Durch diese Integration wird eine einheitliche Datenstruktur geschaffen, die den Anforderungen der internen Analyseprozesse von Daimler Truck entspricht.

# Protokollierung und Fehlerbehandlung

## Log-Dateien und Fehlerprotokollierung

Im Verlauf der Verarbeitung werden alle wichtigen Bearbeitungsschritte und Fehler in Log-Dateien festgehalten. Diese Log-Dateien dienen der Nachverfolgbarkeit des gesamten Verarbeitungsprozesses und ermöglichen eine effiziente Fehleranalyse.

Jede Log-Dateien enthält detaillierte Informationen zu folgenden Aspekten:

* Verarbeitungsfortschritt: Der Verarbeitungsfortschritt jedes Fahrzeugs wird dokumentiert. Sowohl die erfolgreich verarbeiteten Daten als auch die aktuellen Verarbeitungsschritte.
* Fehlerprotokolle: Fehlerhafte Dateien oder unerwartete Probleme werden sofort in einem Fehlerprotokoll erfasst. Die Ursachen für die Fehler (z. B. ungültige Dateiformate oder unvollständige Daten) werden ebenfalls dokumentiert.

Diese umfassende Fehlerdokumentation ermöglicht das zügige Erkennen von Problemen und gegebenenfalls die Durchführung manueller Korrekturen. Zusätzlich wird die Möglichkeit zur weiteren Analyse und Verbesserung der Verarbeitungsprozesse eröffnet.

## Error.Log: Struktur und Inhalte

Die Error**.Log**-Datei dient zur Fehlerdiagnose und Nachverfolgbarkeit innerhalb der Merger-Applikation. Sie dokumentiert detailliert alle Fehler, sodass technische Probleme effizient bearbeitet werden können.

Ein typischer Aufbau der Error.Log enthält folgende Informationen:

Allgemeine Angaben:

* Datum und Uhrzeit: Der Zeitpunkt, an dem der Fehler aufgetreten ist.
* Konfiguration: Die Art der betroffenen Datei, z. B. Loggerdaten oder SOME\_IP-Daten.
* Fahrzeug-ID: Eindeutige Identifikation des betroffenen Fahrzeugs.

Datei- und Verzeichnisinformationen:

* Import-Root: Das Verzeichnis, aus dem die Datei geladen wurden.
* Import-Datei: Die Merge-Datei, bei deren Schreibevorgang der Fehler aufgetreten ist.

Fehlerbeschreibung:

* Art des Fehlers: Detaillierte Angabe des Problems, z. B. „Permission denied“, „Ungültiges Dateiformat“ oder „Unhandled Exception“.
* Folgen: Die Konsequenz des Errors, z. B. „Konfiguration wurde deaktiviert“.

## Einfluss fehlerhafter Dateien auf den Gesamtprozess

Fehlerhafte Dateien werden isoliert behandelt, sodass die Verarbeitung der Daten anderer Fahrzeuge nicht gestoppt werden muss.

Vorgehen:

* Fehlererkennung: Die Art des Fehlers wird erkannt (z. B. falsches Format, Zugriffsprobleme).
* Prozessabbruch: Die Verarbeitung des betroffenen Fahrzeugs wird gestoppt und die Konfiguration deaktiviert.
* Protokollierung: Die Fehlerdetails werden in der **Error.Log** dokumentiert, sodass man genau einsehen kann, was bei dem betroffenen Fahrzeug der Fehler war.

Dieser Ansatz sichert die Stabilität der Applikation und erleichtert die Fehlerbehebung ohne, dass der Gesamtprozesses darunter leidet.

# Auswertung (Statistik)

Die Inhalte dieses Kapitels basieren auf den internen Unterlagen von Daimler Wörth (2013).

## Analyse und Auswertung der verarbeiteten Daten

Die Auswertung der verarbeiteten Daten erfolgt durch die Erstellung von Berichten, die wesentliche Kennzahlen und Parameter der Fahrzeuginformationen übersichtlich und grafisch darstellen. Die Auswertung umfasst folgende Schwerpunkte:

* Zeitliche Darstellung der Daten:
Es wird eine grafische Darstellung von Parametern wie Fahrzeuggeschwindigkeit, Motordrehmoment und Querbeschleunigung im Verlauf der aufgezeichneten Zeit erstellt. Durch die grafische Darstellung kann man sofort erkennen, wann das Fahrzeug zum Beispiel stand, oder zu welchem Zustand des Fahrzeugs welche Temperatur gemessen wurde.
* Statistische Kennzahlen:
Die Berechnung von Minimum, Maximum, Mittelwert und weiteren statistischen Werten für jeden Messkanal, z. B.:
	+ Fahrzeuggeschwindigkeit: 36,4 km/h (Mittelwert), 82,2 km/h (Maximum)
	+ Motordrehmoment: 307,5 Nm (Mittelwert), 2198 Nm (Maximum)
* Fahrzeit und Standzeit:
Es wird eine Analyse der Fahrzeuginaktivität (Standzeiten) und der aktiven Fahrzeit pro Schicht erstellt.
* Parameterübersicht:
Physikalische Größen wie Kühlwassertemperatur, Umgebungslufttemperatur oder Fahrzeugmasse werden zusammengefasst, um Betriebszustände des Fahrzeugs zu bewerten.

Die Ergebnisse werden in Tabellen und verschiedenen Grafiken dargestellt, um Trends und Anomalien leicht erkennbar zu machen. Dadurch wird die Auswertung der Daten für Daimler Truck deutlich erleichtert, da sie sich direkt die Ergebnisse in Grafiken anschauen können, mit denen sie direkt erkennen können, welche Daten ihnen passen und an welchen Stellen sie eventuell am Bau des nächsten Fahrzeugs noch Änderungen vornehmen sollten.

## Statistische Verfahren zur Überprüfung der Datenqualität

Durch die Statistiken in der Auswertung kann die Datenqualität überprüft werden. Dadurch erkennt man schnell fehlerhafte oder auffällige Daten und die Zuverlässigkeit der Ergebnisse wird sichergestellt.

Zentrale statistische Verfahren:

* Berechnung von Minima, Maxima und Mittelwerten:
Für jeden Messkanal werden die Extremwerte und der Durchschnittswert berechnet, wodurch ungewöhnliche Daten sofort auffallen.
	+ Beispiel: Eine ungewöhnlich hohe Kühlwassertemperatur kann auf einen Sensorfehler oder ein Problem im Kühlsystem hinweisen.
* Verteilung und Perzentile:
Die relative und kumulative Häufigkeit der Messwerte wird ebenfalls aufgegriffen. Die Berechnung von Perzentilen (z. B. 10., 50. und 90. Perzentil) gibt einen Überblick über typische und extreme Betriebsbedingungen.
* Trend- und Anomalieerkennung:
Durch Zeitreihenanalysen werden zeitabhängige Muster leicht erkannt. Abweichungen vom typischen Verlauf (z. B. plötzlicher Geschwindigkeitseinbruch) können auf Messfehler oder unvorhergesehene Ereignisse hindeuten.
* Ausreißeranalyse:
Wenn ungewöhnlich hohe oder niedrige Werte werden identifiziert werden, werden diese separat überprüft, um deren Ursache zu klären.
* GPS-Koordinaten: Die Koordinaten werden auf einer Karte visualisiert, um Auskunft über die gefahrene Strecke zu geben.

## Grafische und tabellarische Darstellung der Ergebnisse

Die Auswertungsergebnisse der Logger- und SOME\_IP-Daten werden in einer Kombination aus Tabellen und Grafiken dargestellt. Dies erleichtert die Analyse, indem sie komplexe Daten anschaulich und leicht verständlich machen.

Tabellarische Darstellung:

* Zusammenfassung der Schlüsselparameter:
Die Tabellen enthalten statistische Kennzahlen wie Minima, Maxima, Mittelwerte und Perzentile für ausgewählte Messkanäle.
	+ Beispiel:
		- Fahrzeuggeschwindigkeit: Min. 0,0 km/h, Max. 82,2 km/h, Mittelwert 36,4 km/h.
		- Motordrehmoment: Min. -1151,2 Nm, Max. 2198,0 Nm, Mittelwert 307,5 Nm.
* Vergleich der Betriebsparameter:
Durch die tabellarisch untereinander Aufzeigenden Darstellung kann man direkt die Extrema verschiedener Parameter vergleichen und zusammenhänge erkennen.

Grafische Darstellung:

* Zeitreihenanalysen:
Verlaufsdiagramme zeigen den zeitlichen Verlauf von Parametern wie Geschwindigkeit, Motordrehmoment und Querbeschleunigung. So können Muster, wie z. B. Beschleunigungs- und Verzögerungsphasen, einfach erkannt werden.
	+ Beispiel: Ein Diagramm der Fahrzeuggeschwindigkeit verdeutlicht Fahr- und Standzeiten im Tagesverlauf.
* Histogramme und Verteilungen:
Histogramme zeigen die Häufigkeitsverteilung von Parametern, z. B. welche Geschwindigkeit am häufigsten gefahren wurde.
	+ Beispiel: Die Summenhäufigkeit der Fahrzeuggeschwindigkeit visualisiert typische Fahrbedingungen.

Die Grafiken sind alle auf den gleichen Zeitraum abgebildet, um sich beim Anschauen leichter die Messwerte erklären zu können. Man sieht sofort, welche Parameter gegenseitigen Einfluss aufeinander haben, wenn zum Beispiel bestimmte Werte zum Gleichen Zeitpunkt einen ähnlichen Anstieg aufzeigen.

# Zusammenfassung und Ausblick

In dieser Arbeit wurde ein System vorgestellt, das die Verarbeitung und Aufbereitung von Logger- und SOME\_IP-Daten bei Daimler Truck vereinfacht und effizienter gestaltet. Mit der Merger-Applikation ist es gelungen, einen zuverlässigen und klar strukturierten Prozess zu schaffen, der große Mengen an Messdaten automatisiert in übersichtliche Tagesdateien umwandelt und eine zeitnahe Bereitstellung der Daten ermöglicht.

Besonders hervorzuheben ist, dass die Daten nun standardisiert und fehlerfrei verarbeitet werden können. Die klare Trennung der Datenformate, die flexible Anpassung durch die INI-Datei und die transparente Fehlerprotokollierung machen die Lösung robust. Die Mitarbeiter\*innen von Daimler Truck profitieren von einer erheblichen Entlastung, da viele manuelle Arbeitsschritte entfallen und Flüchtigkeitsfehler reduziert werden.

Auch die Auswertung der Daten hat sich durch die Anwendung spürbar verbessert. Statistische Analysen und grafische Darstellungen liefern ein klares Bild von Betriebsparametern, Trends und Auffälligkeiten. Dadurch wird nicht nur die Datenqualität erhöht, sondern auch die Grundlage für Entscheidungen in der Fahrzeugentwicklung gestärkt.

Ausblick:
Die Merger- Applikation hat das Potenzial, in Zukunft noch viele Erweiterungen zu bekommen:

* Bessere Fehlerprüfung: Mit weiteren Plausibilitätschecks könnten Fehler frühzeitig erkannt und automatisch behoben werden.
* Einfachere Bedienung: Eine benutzerfreundliche Oberfläche würde den Umgang mit der Applikation noch einfacher und übersichtlicher gestalten.
* Effizientere Datenspeicherung: Neue Verfahren zur Datenkomprimierung könnten den Speicherplatzbedarf senken.

Durch die Merger-Applikation wurde ein erster, wichtiger Schritt gemacht, um die Verarbeitung von Messdaten bei Daimler Truck deutlich zu verbessern und zu automatisieren. Sie bietet eine stabile Grundlage, auf der weitere Verbesserungen und Erweiterungen aufbauen können, um den weiteren Anforderungen im Bereich der Fahrzeugentwicklung gerecht zu werden.

# Literaturverzeichnis

1. *Daimler Wörth (2024). Internes Lastenheft: 2024-06-25 Lastenheft\_Merger. Unveröffentlichtes internes Dokument.*
2. *measX GmbH & Co. KG (2024). Interne* ***Anforderungsbeschreibung****: DaimlerTruck\_MBTaps\_Datenprozessierung MF4 Dateien Basisapp Teil 1 und 2. Unveröffentlichtes internes Dokument.*
3. *Daimler Wörth (2013). Beispiel einer Auswertung: Triebstrang. Unveröffentlichtes internes Dokument.*