Game changer connectivity:
Remote analytics as the key to new findings in vehicle development and product maintenance
Connected vehicles offer OEMs, Tier 1 suppliers, and other parties new possibilities for diagnosis and data analysis. Instead of being dependent on data that is handled in the repair shop, case-related questions can be answered by the targeted acquisition of current fleet data. The software specialist Elektrobit has developed a complete workflow to collect and analyze this type of vehicle data.

The traditional approach is just as rudimentary as it is unsatisfactory: Log data and error logs of electronic control units are read and diagnosed in the repair shop and provided to the OEM where required. Consequently, these data are not available in real time. Besides, they are only collected if the driver has a problem with his vehicle or visits his dealer for a routine inspection. This process can be considerably improved by remote diagnostics over existing vehicle networking. From that, OEMs, users, and vehicle owners can benefit equally. Even for dealers and repair shops this approach offers interesting business models. With its remote analytics system EB cadian, Elektrobit provides a complete solution based on a workflow that it has developed itself for applications like this.

Figure 1: Specific targeted data collection with EB cadian

- Data efficiency to save bandwidth
- Principle of data economy for compliance/privacy
The solution adopts an on-demand approach to acquiring and analyzing vehicle data. In contrast, there are also systems on the market that continuously collect a set of measurements - or vehicle data, respectively - with a specific acquisition rate. These data are provided for subsequent analysis. This approach may be well suited for deriving comprehensive findings on the behavior of a complete fleet of vehicles, but it nonetheless has disadvantages too: Only pre-defined data are available for analysis. The load for mobile transfer and the back-end systems used for storage is considerable. And in some markets, including the data-protection sensitive German market, a process of this type poses non-negligible legal problems.

On the other hand, a case-related, on-demand approach for data analysis offers clear advantages: It supports special issues for which very specific vehicle data has to be surveyed and analyzed. It works in a sustainable and considered way on the principle of surveying data as economically as possible. Analysts, though must precisely specify the issues that they are investigating beforehand. EB cadian adopts this approach.

**Workflow defined for remote analytics**

In a data analysis that is carried out this way, the typical workflow includes formulating exact questions, defining the data and the vehicles to be considered, acquiring data, carrying out quality assurance of the global data pool and the analysis based on this as well as processing the results. The prepared reports and, where required, the data exports are then the basis for further process steps. This process can best be illustrated by a practical use case: An analysis of possible starter battery malfunctions in winter.

The questions formulated for an analysis of this type could for example be: "Are there starter batteries in the vehicle fleet that behave significantly differently from most other batteries?" The assumption here is that a behavioral outlier like this represents an indication of a potential malfunction. In order to be able to answer this question on the basis of quantifiable data, the "record battery voltage level 900 times every 1000 milliseconds" task can be construed as an example of this. In addition, this requirement would typically be limited to a part of the vehicle fleet that is particularly relevant for the analysis; in the example mentioned, these vehicles could be those with gasoline motors that are more than three years old. EB cadian defines a so-called domain-specific language (DSL) for requirements and specifications like this. A DSL is a computer-interpretable language that has been specifically developed for a special focus. In the case of EB cadian, this is the targeted collection of vehicle data by analysts. In specifying the data collection task, it is important that it remains agnostic with respect to individual vehicles: It must not contain any details on vehicle-internal data production and representation. The vehicle is then selected from the total fleet on the basis of properties such as type of fuel, mileage, model or equipment line, and similar characteristics. As far as the system is used in an OEM that has access to specific data such as production date, chassis number, serial number or part number, and similar information, these data can also be used as filter characteristics.

Specific data acquisition is derived from the experiment definition which is formulated in this DSL and limited to relevant fleet vehicles. The EB cadian analysis platform creates vehicle-specific information from the vehicle-agnostic representation and assigns it to the individual vehicles that are determined by the filter. The appropriate electronic control units only carry this out in targeted vehicles. Part of the data is processed in the automobile, for example standardization and conversion of measurements to SI units or the simple linking of several measurements to a derived quantity. Different data sources are considered here: information from the CAN bus, diagnostics data, or the driver’s operations such as touch-triggered HMI events.
The next important step is quality assurance. Time and again, individual measurements will be erroneous: for example lying outside of a plausible range of values, having damaged time stamps, or other issues. Data consistency should therefore be tested and ensured before the analysis. Besides filtering out implausible values, this can also mean interpolation of missing measurements. Depending on the task, a data set may also have to be excluded from further processing. This could be the case, if it contains more than a given percentage of measurements that have been detected as potentially incorrect. Subsequently, the erroneous measurement set has to be “thrown away”.

Then, the data are analyzed according to the originally formulated problem type. As in the given example of the starter battery, this can be an identification of outliers or anomalies: the time series of each vehicle is converted into a histogram in order to establish comparability between vehicles. In the next step, a so-called distance function that measures voltage histogram similarity is analyzed between two vehicles at a time in pairs. This comparison can be made between the participating vehicles in a single data collection process or even through several runs. After all comparison pairs have been analyzed, vehicles that have only a few or no “neighbors” compared with the voltage histogram (“k-nearest neighbor”) can be identified. This approach, which was kept deliberately simple in the previous example, can be extended to multi-variant data, application-specific transformations of the time axis, or other pattern-recognition algorithms. An advantage of this method is that it represents a so-called unsupervised method of machine learning. No expensive training data is required to carry this out. Of course, more complex algorithms, that are designed in a problem-specific way and can include additional information in the analysis process, are also conceivable at this point.

Figure 2: EB cadian workflow

The encrypted and signed diagnostics data from the electronic control units are transferred to the back-end to be collected there. In the battery example above, this would then be the voltage progression of the battery over the defined acquisition period.

Quality assurance and analysis techniques

The next important step is quality assurance. Time and again, individual measurements will be erroneous: for example lying outside of a plausible range of values, having damaged time stamps, or other issues. Data consistency should therefore be tested and ensured before the analysis. Besides filtering out implausible values, this can also mean interpolation of missing measurements. Depending on the task, a data set may also have to be excluded from further processing. This could be the case, if it contains more than a given percentage of measurements that have been detected as potentially incorrect. Subsequently, the erroneous measurement set has to be “thrown away”.

Then, the data are analyzed according to the originally formulated problem type. As in the given example of the starter battery, this can be an identification of outliers or anomalies: the time series of each vehicle is converted into a histogram in order to establish comparability between vehicles. In the next step, a so-called distance function that measures voltage histogram similarity is analyzed between two vehicles at a time in pairs. This comparison can be made between the participating vehicles in a single data collection process or even through several runs. After all comparison pairs have been analyzed, vehicles that have only a few or no “neighbors” compared with the voltage histogram (“k-nearest neighbor”) can be identified. This approach, which was kept deliberately simple in the previous example, can be extended to multi-variant data, application-specific transformations of the time axis, or other pattern-recognition algorithms. An advantage of this method is that it represents a so-called unsupervised method of machine learning. No expensive training data is required to carry this out. Of course, more complex algorithms, that are designed in a problem-specific way and can include additional information in the analysis process, are also conceivable at this point.
In the end, the reported results are processed in the form of text or graphic data or even as numeric values, for example, in the form of CSV files in order to process the analysis results externally as required.

**Figure 3: EB cadian deviation**

The result is, of course, primarily available for the person who has defined and ordered the original analysis task. Analysis experts are, however, typically disseminators in their companies who share specific processes, results, and findings with other parties as well as allow occasional users of the system to process complex tasks. The consequences that arise from an appropriate analysis – partially automated or, where required, entirely automated – could, for example, lead to a notification in the vehicle. In the battery example, a display message that indicates a potentially soon occurring battery malfunction would be considered. In turn, additional business models could be based on this like a special promotion for installing a new battery.

**Focus on data protection and data security**

Data protection and data security are important aspects that have played a key role in designing EB cadian. Data collected as part of remote analytics are in any case anonymized. Both, front-end access by the analysts and system access to each individual vehicle require mutual authentication. All connections and all transferred data sets are encrypted according to strong cryptographic processes. And all queries and data communication processes are documented with respect to the compliance guidelines that may apply, and to audits that may be required in encrypted and access-protected log files as well.

Elektrobit has defined best practices for using its system, and for the security of cloud applications: A risk and threat assessment, which helps to identify the necessary safety concepts and elements, should be carried out, especially with regard to data protection and data integrity. As part of software developments, code quality should be monitored and ensured by using a quality management platform. It is principally advisable to automate as many parts of the processes as possible regarding development, provisioning, testing, roll-out, and monitoring. Reproducibility of the results, their reliability, and the possible speed of implementation clearly increase this way.
Besides security questions, scalability of the platform and systems that are included by OEMs and Tier 1 suppliers are also an important factor. With its stateless micro-service architecture, EB cadian is state of the art. For its own developments and extensions, Elektrobit advocates the “think big” principle because it is generally very hard to scale a system that has not been designed from the outset on the basis of horizontal scalability for very heavy loads. The usual way to simply buy bigger machines quickly leads to a dead end. Only an architecture that is scalable from the outset can achieve this. Another aspect of scalability is the reverse case: At times of low loads, the system should make it possible to free up unused resources and save costs this way.

**Figure 4: Advantages scalable Cloud**

In order to support the broadest possible range of applications, EB cadian is modular in construction and can be integrated in existing fleet and model management systems or analysis tools. By using on-demand data analytics in the field, OEMs and Tier 1 companies open up many new possibilities to improve their development processes and products as well as to create new services and business models.
Dr. Oliver Pajonk

is a Connected Car Expert at Elektrobit Automotive GmbH (EB) and holds a doctoral degree in Numerical Mathematics and Scientific Computing from the Technical University Braunschweig. He has over 10 years of experience in the oil and gas industry, creating exploration and planning software products with the main focus on large scale uncertainty quantification and optimization. He joined EB in 2015 where he is now focusing on remote diagnostics and big data analytics.

Sebastian Fräßdorf

is Project Manager Connected Car at EB. He has a background in Electrical Engineering and developed software for engine control units. In 2007 he realigned as a system engineer with focus in project-, requirements- and test management. He joined EB in 2012, after earning his Master in Business Engineering.

Axel von Engel

studied Computer and Communications Systems Engineering at the TU Braunschweig. After graduating in 2010, he worked on a software product for timing analysis at a company specialized on analysis of real-time systems. He joined EB in 2013 where he develops enterprise and cloud software solutions.
About EB Automotive

Elektrobit (EB) is an award-winning and visionary global supplier of embedded software solutions and services for the automotive industry. A leader in automotive software with over 25 years serving the industry, EB’s software powers over 70 million vehicles and offers flexible, innovative solutions for connected car infrastructure, human machine interface (HMI) technologies, navigation, driver assistance, electronic control units (ECUs), and software engineering services. EB is a wholly owned, independent subsidiary of Continental AG.

---

Game changer connectivity: Remote analytics as the key to new findings in vehicle development and product maintenance