

# PRO-OPT

Collaborative Data Analysis in  
the Automotive Industry



# The Challenge

Modern industrial production processes are characterized by a high degree of division of labor. In addition to the physical products themselves, the huge amounts of data emerging from the production steps are becoming increasingly important. This also increases the significance of **digital cooperation** and networking among individual companies in decentralized data systems – so-called **Smart Ecosystems**. This development poses great challenges for industry

## The PRO-OPT Application Scenario

PRO-OPT uses the **automotive industry** as an example for such ecosystems. Vehicles constitute extremely complex systems of systems consisting of components from a wide variety of suppliers. This complexity is also reflected in the associated ecosystem. The automotive manufacturer is preceded by a multi-step supply chain. The production of a vehicle is followed by the remainder of the life-cycle, manifested, e. g., in the form of visits to the repair shop or also in the form of connected-car solutions.

If systematic problems arise in the product life-cycle, these must be analyzed and eliminated as quickly as



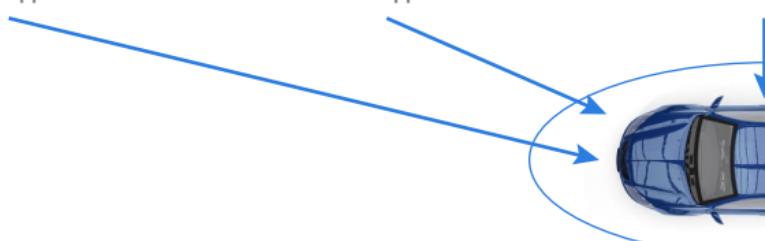
Parts supplier

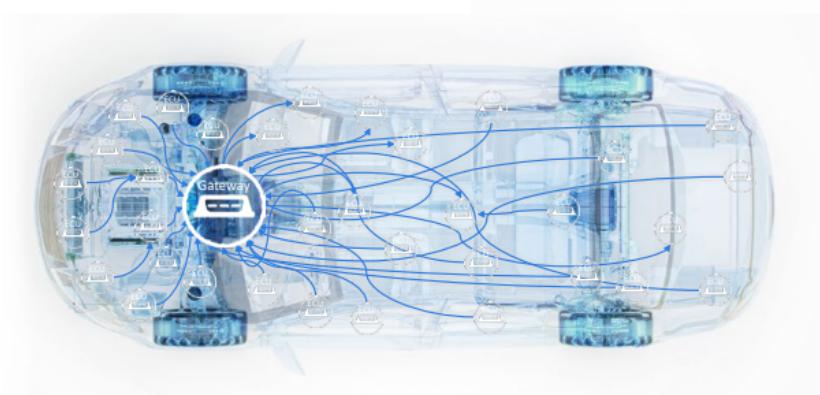


ECU supplier

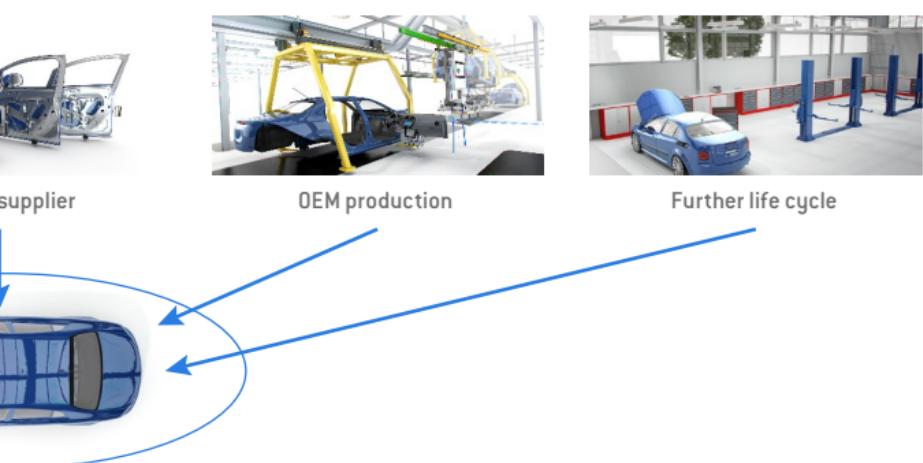


Module





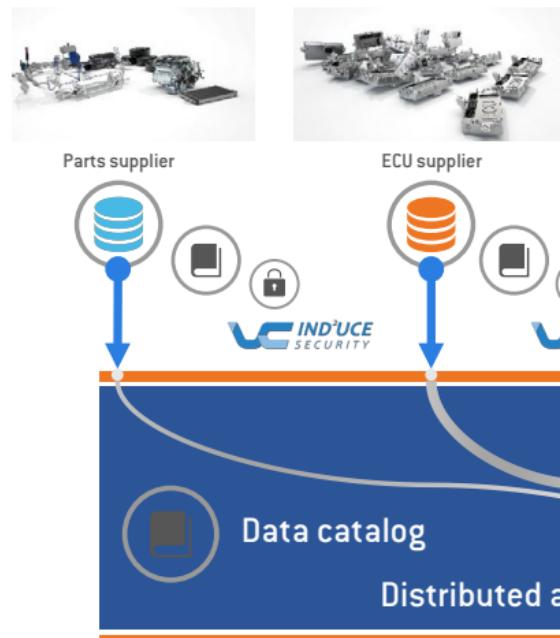
possible. **Smart Data analyses** here represent an ideal instrument, for example for root-cause analyses or in the case of traceability considerations. They require data from all participants in the ecosystem. These data are available in heterogeneous source systems in varying degree of quality. Ideally, one would aggregate them in one central place and evaluate them there. In practice, this is not desired due to the economic interests and the protection of the intellectual property of the companies involved. It must be possible to integrate in a controlled manner only such data into an overarching analysis that are needed in the respective case, and to always take into account the **protection of the local interests** when doing so.



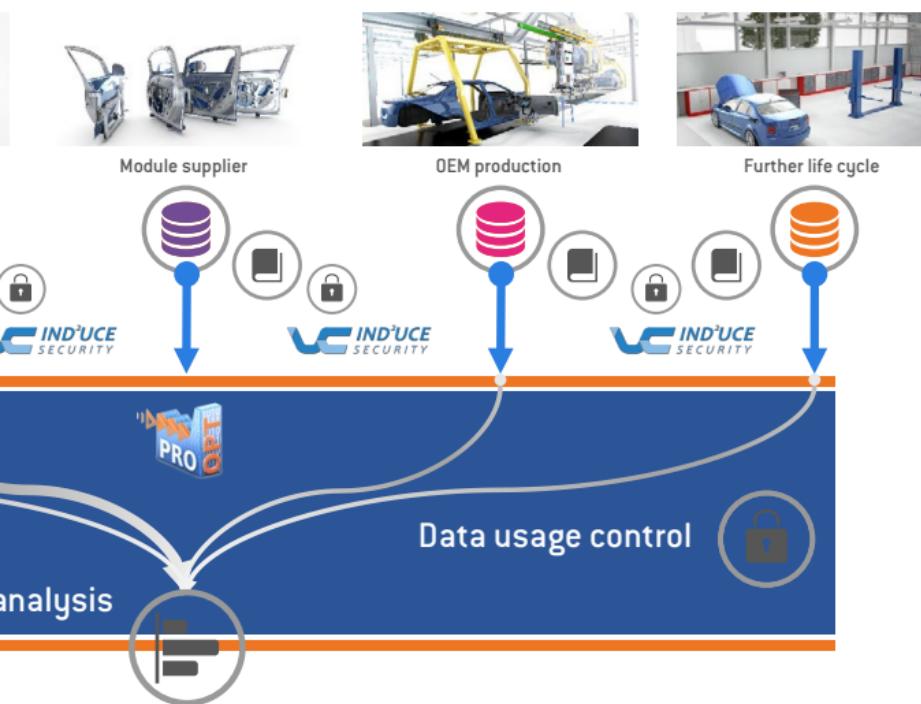
# The PRO-OPT Platform

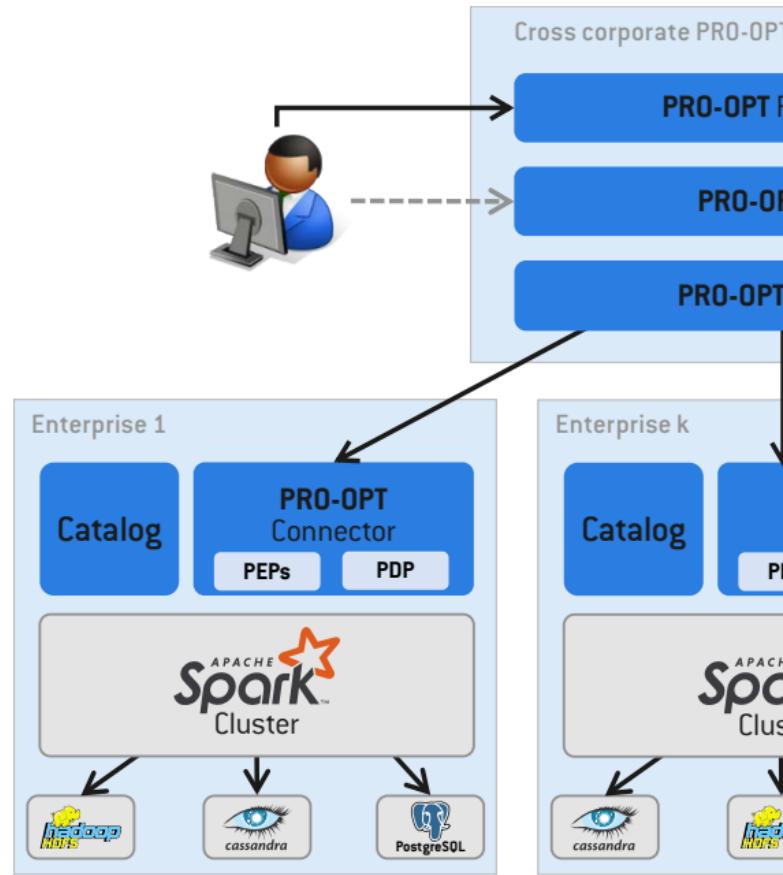
This is where the PRO-OPT platform comes into play. The PRO-OPT platform enables comprehensive Smart Data analyses in complex ecosystems for specific purposes. The underlying principle is that the raw data never leave the sovereignty of the individual company. Instead, **comprehensive analyses** are requested centrally from the platform; these are then divided into sub-analyses, which are performed locally on the individual data sources. The resulting individual findings are then aggregated in the platform into an overall result. The platform performs the distribution of requests and feedback of the results, offers interfaces for **connecting existing heterogeneous data sources**, and provides support for the **fusion of the individual results** with the help of a graphical programming language.

In order to be able to request analyses, the respective analysis possibilities that are available and the meaning of the underlying data from the individual data sources must be known. To this end, the participants in the ecosystem can describe their data sources and the associated analyses in the form of a **data catalog**. The released portions of the data catalogs can be accessed centrally on the platform. The data catalogs also describe **quality features** of the analyses offered, resp. of the underlying data. The quality information is shown



to the users of the platform and can be taken into account during the fusion of the individual results. The **data usage constraints** are also defined in the catalog. Beyond mere access rights, these allow detailed control over which analyses are permitted on which data in which specific case. Analyses are always **linked to a specific purpose**. An OEM, for instance, can limit a supplier's analyses to the vehicle data of such vehicles for which a joint test is currently being performed. The frequency of analyses as well as the granularity and size of the result set can also be controlled. For every analysis, the local **data usage control** components of the PRO-OPT platform check compliance with the defined constraints, i.e., they permit or deny accesses and, if requested, anonymize or pseudonymize the result data before these leave the company. In this way, the individual interests are protected at all times.

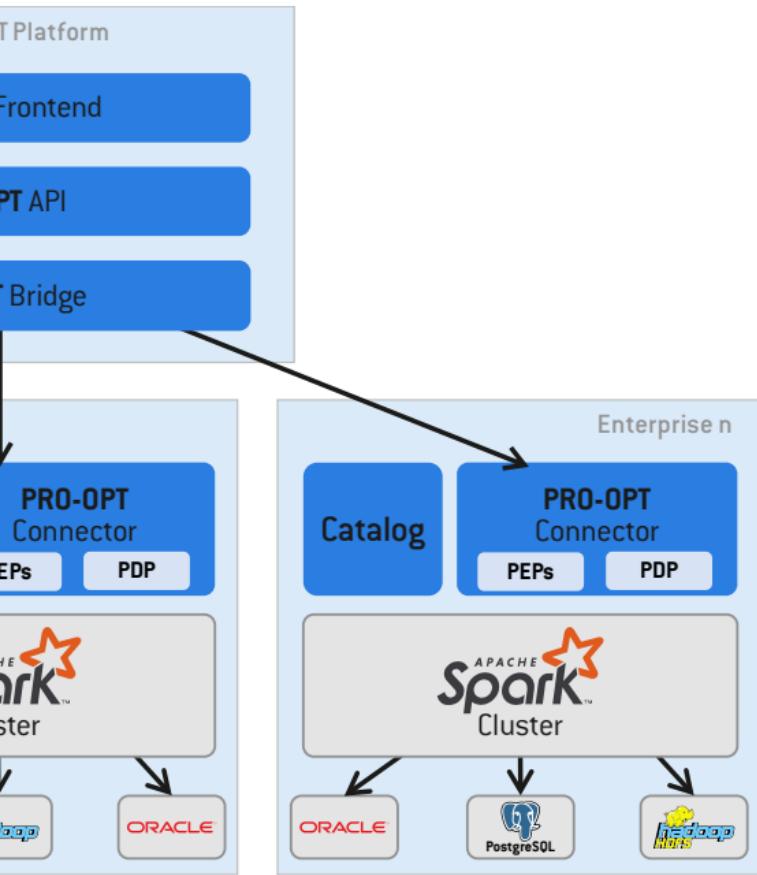




## The Implementation

The PRO-OPT platform consists of **central and local company components**.

PRO-OPT does not invent new basic technologies. The goal is rather to **use existing technology in an advantageous manner**. This means, e.g., that Apache Spark™ is used as the basis for the comprehensive integration, the connection to data sources, as well as for the performance of analyses by means of the Spark-MLlib. The ETL tool Kettle and the associated editor Spoon are used for data fusion. Data usage control is being implemented via the Fraunhofer IESE framework IND²UCE. For data storage at the companies, existing storage solutions are used; for example, Cassandra is used in the application scenario as a database for the optimized storage of traceability data.



## Contact

[info@pro-opt.org](mailto:info@pro-opt.org)  
[www.pro-opt.org](http://www.pro-opt.org)

## The Project

PRO-OPT is being funded as part of the Smart Data technology program of the German Federal Ministry of Economic Affairs..

Gefördert durch:



aufgrund eines Beschlusses  
des Deutschen Bundestages

# Consortium partners

**Consortium leader, technology and application consultant for vehicle diagnostics**  
DSA Daten- und Systemtechnik GmbH  
Germany  
[www.dsa.de](http://www.dsa.de)



**Technical project management, research partner for data usage control, data quality, architecture, data generation**  
Fraunhofer Institute for Experimental Software Engineering IESE  
[www.iese.fraunhofer.de](http://www.iese.fraunhofer.de)



**Technology and application partner for production systems**  
camLine GmbH  
[www.camLine.com](http://www.camLine.com)



**Application partner for vehicle diagnosis**  
Audi AG  
[www.audi.de](http://www.audi.de)



**Research Partner for Data Mining, integration of system components**  
German Research Center for Artificial Intelligence  
[www.dfki.de](http://www.dfki.de)



# Associated partners

**Research partner for visualization**  
Technische Universität Kaiserslautern  
Lehrstuhl für Computergrafik und HCI  
[hci.uni-kl.de](http://hci.uni-kl.de)



**Data provider and evaluation partner**  
Continental AG  
[www.continental-corporation.com](http://www.continental-corporation.com)



**Multiplier**  
Technologie-Initiative SmartFactory KL e.V.  
[www.smartfactory.de](http://www.smartfactory.de)

