Google Cloud and Analysis of Realtime Process Data

R. Langmann
Duesseldorf University of Applied Sciences/Competence Center Automation Duesseldorf (CCAD), Duesseldorf, Germany, langmann@ccad.eu

Abstract: *A current topic regarding the automation of technical processes involves use of the technology trend Big Data. Previous Big Data solutions are concentrated on process data that is not time critical from the upper automation levels for improving coordination and scheduling tasks. In contrast, this paper presents the concept and implementation for a Big Data automation service which can be used to store real time process data from systems/machines in a cloud, from where it can be analyzed with user-specific algorithms.*

Keywords: *Web-based automation system, Big Data analysis, realtime process data, Big Data services, pro-active maintenance.*

I. INTRODUCTION

A current topic for the automation of technical processes in conjunction with the implementation recommendations Industry 4.0 [1] concerns the use of the technology trend *Big Data* and the question of data management for this data. The following central challenge is associated with this: In the Internet of things and services as well as in the planned CPPS systems in the factory of the future (CPPS – Cyber Physical Production System) enormous data volumes are set to result. Whoever knows how to successfully collect, manage and utilize this data is sure to have competitive advantages. This statement concerns Industry 4.0 with the extensive data volumes from production and logistics to a very special extent.

Outside the world of automation, Big Data technologies are already used for different purposes. At the same time, this technology makes it possible to store and analyse immense volumes of data comprising several hundred terabytes and even petabytes using available cloud and server platforms such as Microsoft Azure, Amazon Web Services, Google Cloud Services. One of the first applications are already known for automation technology e.g. analysis of process data of a wind farm with Microsoft Azure [2].

Previous Big Data solutions are concentrated on process data that is not time critical from the upper automation levels for improving coordination and scheduling tasks. Realtime process signals res. realtime processing (RTP) data from the field level (sensor and actuator signals) are not considered for this. However, a large amount of data results here. It is undisputed in technical circles that the analysis of RTP data has a high level of potential for the improving the operating efficiency of machines and systems in future [3]. However, suitable architectures and interfaces as lacking in the prior art in order to be able to provide customized Big Data solutions as automation services from a cloud.

As part of the preliminary investigation regarding future research projects in the Competence Center Automation Düsseldorf (CCAD), the principle feasibility of a Big Data analysis for RTP data[1] should be examined and an initial prototypical solution implemented from this. The prototype solution should be based on the results of the research project WOAS (Web-Oriented Automation System [4]), using a fast Web connector developed in this project for transferring the process data [5] and providing the required Big Data functions as automation services. The following tasks were to be solved in individual detail:

- Examination of cloud system available on the market in regard to their suitability for storing and analyzing real-time process data and selecting a cloud system.
- Development of a prototype solution for Big Data storage in the cloud.
- Development of a prototype solution for Big Data analysis of the stored data.
- Provision of the Big Data functions for storing and analysis as automation services in the WOAS integration portal.
- Testing and evaluating the solutions at an automation system as a demonstrator.

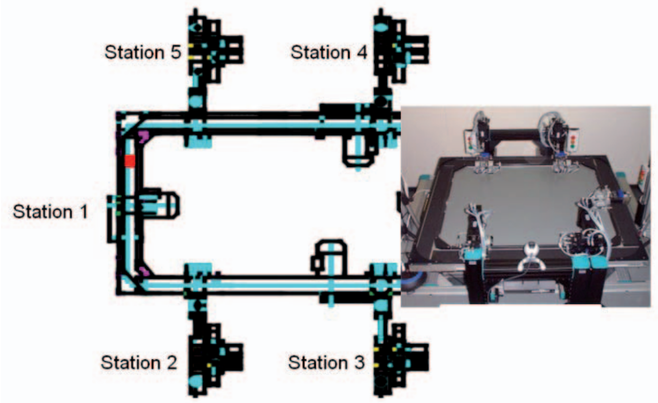

Fig. 1. Assembly/Disassembly system for model cars as a demonstration system (demonstrator) for the Big Data analysis of RTP data.

An assembly system for model cars with five assembly/disassembly stations (Figure 1) functions as a demonstrator for the Big Data solution. The individual stations are operated by PLC controls of various manufacturers. Every station has pneumatically actuated elements. The assembly parts are fed to the individual stations via a workpiece holder on an electrically operated transport conveyor. The system contains a total of 132 digital process signals. This includes

[1] Realtime refers here to process data that is available via an IP network with cycle times of about 50 … 200 ms.

actuator signals for switching the pneumatic valves as well as sensor signals that detect the end positions of the pneumatic valves. The demonstrator system is equipped with standard industrial components and compatible with a real production system.

According to [6], cloud computing is a model that allows a shared pool of configurable computer resources (e.g. networks, servers, storage systems, applications and services) to be conveniently accessed at any time if required. These resources can be provided rapidly and with minimum management expense or low service provider interaction. Cloud computing enables the realization of computer and memory intensive tasks on the infrastructure of an external service provider. In this way, it is possible to access almost unlimited resources without maintenance and service costs. That means cloud services are especially suitable for storing and analyzing data volumes (Big Data - BD).

Access to the cloud services is generally via a RESTful API[2]. It is therefore possible to access the externally stored data or resources via local applications. Depending on the provider, there may be differences in the scope and variety of the supporting programming languages. The services are generally calculated per use. Depending on the provider, package prices can also be agreed.

Four cloud services of different providers (Google, Amazon, Microsoft, Splunk) were examined and evaluated in respect to specific requirements for the present task. One of the most important requirements was the availability of a JavaScript API in order to realize the storing and analysis of the data via client-based Web technology (Rich Client). A detailed analysis of this can be found in [7].

The following general statements can be made in respect to the cloud services of the above providers, taking the task at hand into account:

- *Google:* Google BigQuery is the most suitable cloud service when it comes to the Big Data analysis of RTP data via Web technologies. In contrast to other Big Data services, BigQuery provides a JavaScript API. It is not therefore necessary to store data on the server side and then transfer it to the cloud service via PHP or Java. The division of the data into tables and not into unsystematic files is also good for the process data logging. This provides the option of using the memory additionally as a backup. If required, the data can then easily be downloaded as a CSV file. The limitation of the upload rate is a disadvantage. This requires the buffering of data in the browser.
- *Amazon:* The Amazon Web Service is the best developed service among the above cloud providers. The services can be utilized flexibly and can be combined with one another. However, complications can also result due to the high number of services. The use of streaming services for the analysis of realtime behaviour is unique in the evaluated cloud systems.
- *Microsoft:* Microsoft Azure is the most economical cloud service with the best structured interface. Azure offers numerous options for utilizing virtual instances and analyzing data. However, the specialization here is on the operation of virtual machines.
- *Splunk:* Splunk is not a conventional cloud service. Rather, the company provides software that makes it possible to analyze locally stored data and gain insights. For companies with their own server infrastructure and many system logs to be evaluated, Splunk is a good option for collecting and evaluating information locally.

As research reveals that only Google BigQuery so far offers a JavaScript API, this service was used for the implementation of a Big Data automation service.

II. CONCEPT

The Big Data automation service is based on the results of the WOAS project [4, 8], in which a services concept was developed for a CPS integration platform (CPS - Cyber Physical System).

A Big Data service (BD service) essentially comprises three components:

- *Virtual device:* A virtual device functions as a Web gateway and provides the process data to the BD service via a Web connector of the automation device event-based in realtime (50 ...100) ms via industrial interfaces.
- *BD store:* The process data is stored in blocks in the Google Cloud via the BD store service.
- *BD analysis:* The BD analysis accepts the evaluation of the stored mass process data via available analysis tools.

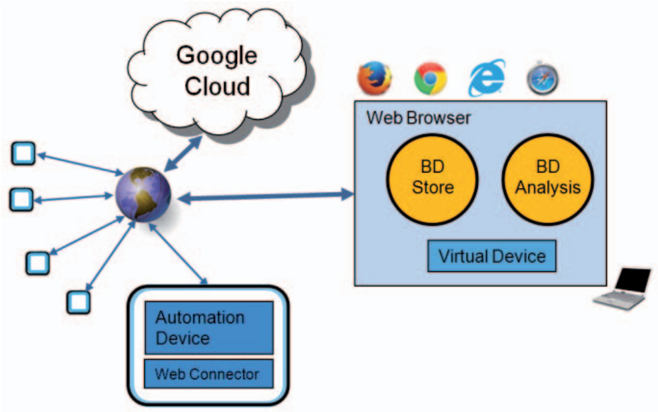

Fig. 2. Schematic diagram of the BD automation services

Figure 2 illustrates the schematic diagram of the BD automation services.

A. Virtual device

The virtual device maps the real automation device into the Web world and offers a uniform access via a lean JavaScript object interface for all automation devices. The Web connector hence enables adaptation of the virtual device to the automation device and functions as a protocol gateway. In terms of a CPS-based automation, the automation device, Web connector and

[2] REST – REpresentational State Transfer, API – Application Programming Interface

virtual device form a CPS component that can be located at any point in the IP network.

B. BD store

BD store is an automation service that is structured according to the WOAS service guidelines [9] and which uses the Google Big Data API for transferring the process data between a Web browser and the Google Cloud. BD store is provided on a Web server as a JavaScript object, loaded in the browser after calling up and dynamically instanced. The communication to the process data is via the virtual device directly via the Web browser without the participation of a server corresponding to Figure 3. BD store sends the incoming process data via POST request in blocks to the Google Cloud.

C. BD analysis

BD analysis enables the analysis and evaluation of the process data stored in the Google Cloud. The service uses the tools provided by Google for this. This includes, in particular, query language similar to SQL, with which complex analyses can also be realized relatively easily. After sending the query command via a POST request, the result is transferred via a response after analysis in the cloud using Google BigQuery. The returned object is processed and visualized via Google charts. The representation is realized as an SVG graphic (Scalable Vector Graphic) and can hence also be displayed in mobile devices. The communication occurs directly between the Web browser and Google Cloud (see Figure 3)

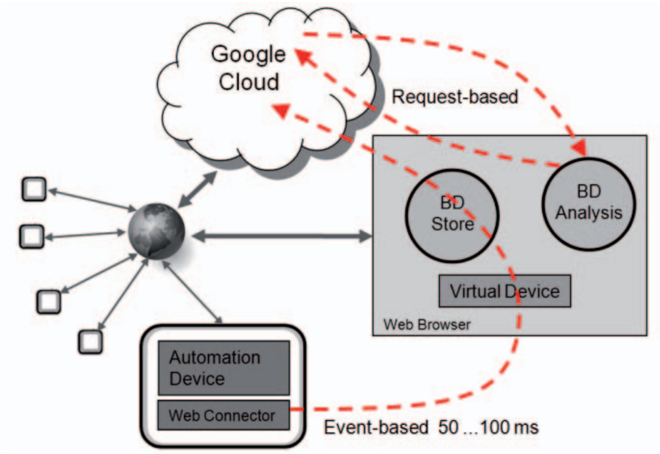

Fig. 3. Process data flow in the BD services system

III. Implementation

The implementation of the two services *BD store* and *BD analysis* was realized as a JavaScript object corresponding to the interface description of an automation service according to the WOAS model [9]. For the transfer of the process data from the automation system used as a demonstrator (see Figure 1) to a Web browser, available Web connectors and virtual devices were used for access to OPC DA drivers, with which the process data can be transferred sufficiently fast at response times of (50 ...100) ms.

Two problems had to be considered for the BD store:

- Limitation in the number of uploads

  Google BigQuery limits the number of daily uploads (memory cycle) to 1000. This means a maximum upload rate of 86.4 s. However, the RTP data is available much more frequently from the system. In extreme cases, new process data is available every 50 ms, which has to be stored in the cloud. The current process data is therefore buffered on the Web page and only saved in blocks via BD store in the Google Cloud. The transfer cycle is currently set to 90 s. This ensures that the maximum transaction rate per day is not exceeded.

- Authentication

  An authentication is required for access to the Google cloud. The authentication is via a Google account, the project instance and a project ID. The authentication via JavaScript can be implemented automatically by using the Google API. Different authorizations can be set easily for various Google services via a user-friendly Web interface. Users who want to use both services BD store and BD analysis require a Google account as a basis, which must be assigned to the relevant Big Data project by an administrator. The further setting of the authentication is then implemented by parameterizing the service instances. The use of a Google service account in conjunction with an application operated in the cloud also enables access to the data independently of the user account. The access monitoring is then at the provider of the service.

Figure 4 shows the representation of the BD store in the Web browser for different states.

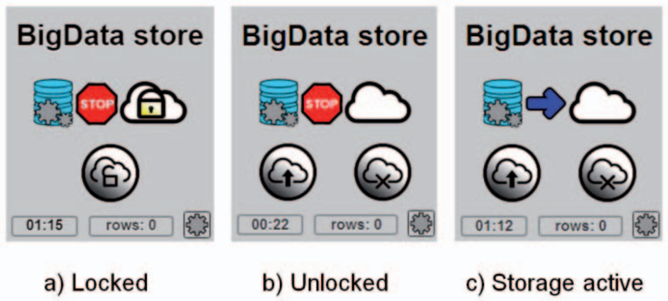

Fig. 4. Operating images of the BD store for different states

A dialogue can be called up for a process data filter via the gear symbol. A BD store instances can store the RTP data of maximum eight automation devices. After activation (Figure 4b), all data transferred to the service is collected and transferred to the cloud after expiry of the block transfer cycle. The field designated "rows" indicates the number of data sets transferred in the last transfer. If the data transfer is terminated using the stop button, the service will consequently stop logging the data. The data collected until then is still transferred after expiry of the current transfer cycle.

The BD analysis serves for query and visualizing the process data stored in the Google Cloud. The service was also compiled as a JavaScript object according the definition of a

WOAS service. The data for the visualization is transferred to the service after successful querying from the cloud, where it is depicted in an SVG graphic via different diagram types from Google Chart. Figure 5 shows the service in a Web browser in unencrypted form for an example query with display as a bar diagram (analysis of the assembly time in relation to the pressure of the compressed air).

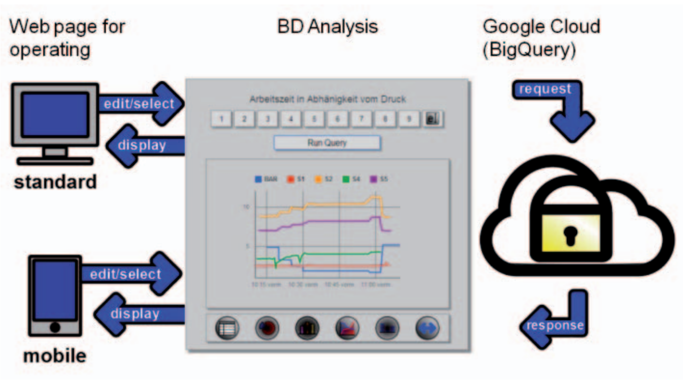

Fig. 5. Analysis and display of an example query with BD analysis in the Web browser

The service is divided into three areas: The upper area of the service serves for selecting and editing the query commands. The query result is visualized in the middle section. In the lower area it is possible to select the required diagram types.

The representation of the query diagrams was implemented with the Charts API provided by Google. The BD analysis supports a series of diagram types (table, bar diagram, line diagram etc.). It is possible to interact in the diagrams by clicking and moving over with the mouse. The line diagram also enables zooming into the representation.

Nine query commands are preset in relation to the demonstrator in the BD analysis service. These include:

- Processing time for the individual stations of the assembly system in relation to the operating pressure of compressed air,
- Number of produced model cars in relation to the days of the week,
- Movement times of the gripper arms at the stations.

The ninth memory slot (Button 9 in Figure 5) only contains one format template for query strings, in which specific queries can be edited by the user. The following source code shows the Google SQL query for determining the most productive time of the day at the system as an example:

```
SELECT IF(HOUR(
MSEC_TO_TIMESTAMP(TIMESTAMP_TO_SEC(time)+7200000))>=
10, STRING(HOUR(MSEC_TO_TIMESTAMP
(TIMESTAMP_TO_SEC(time)+7200000))),'0'+STRING(HOUR(
MSEC_TO_TIMESTAMP(TIMESTAMP_TO_SEC(time)+7200000)))
)+':00 Uhr'as hour_of_day,
SUM(IF(pdname='S7:[S7Verbindung_1]I_WOAS5_AssemblySR
eady' AND value_b=false,1,0) )as number_of_
transporters,FROM[WOAS_D.WOAS]WHERE time >
SEC_TO_TIMESTAMP(NOW()/1000 - 15768000000)GROUP BY
hour_of_day ORDER By hour_of_day;
```

IV. Application

Two different applications were implemented for both services, BD store and BD analysis:

A. *Analysis dashboard*

This application comprises a Web page (see Figure 6), which can perform a cloud-based storing and analysis of RTP data independently of the WOAS integration portal. The Web page makes it possible to have the standard functional scope of the services and provides additional services. Manual inputs can also be made via an special entry range so as to be able to log process signals that are technically not accessible. Additional query commands can also be transferred to the BD analysis service.

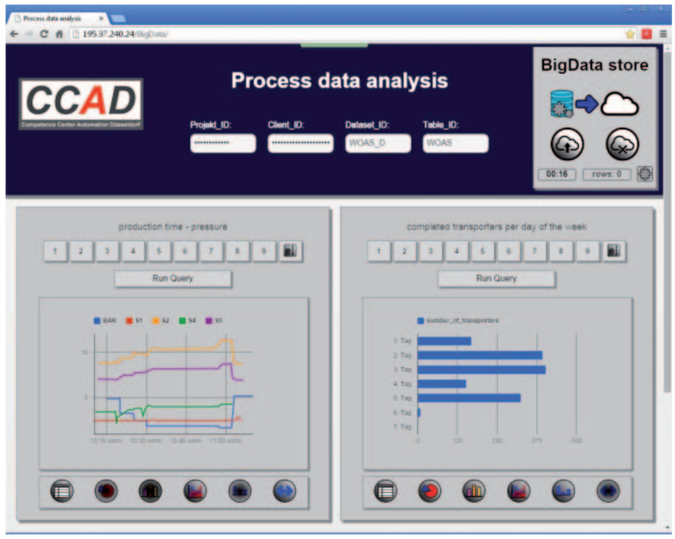

Fig. 6. Analysis dashboard with two instances of the BD analysis service

In addition to the standard Web page, a mobile Internet page was created, this making it possible to enter the manual data and store this in the cloud. The addressing of the desired table is similar to the standard Web page.

B. *BD automation service in the WOAS portal*

The BD automation service was integrated in the WOAS platform in a further application as a prototype service and enables the parameterizable storing of all sensor and actuator signals from the demonstration system in the Google Cloud and their user-specific evaluation using BD analysis. Both services can be freely configured by the user in the WOAS portal and used for a Web-based automation system together with other services. The WOAS portal can be accessed publicly via `http://woas.ccad.eu` The two services can be used in EDIT mode from the WOAS service directory under the designation `service4bdstore` (BD-Store) and `service4bddisplay` (BD analysis).

C. *Test results*

The process data of the demonstrator has been saved in the Google Cloud since April 2014 as part of a test application. Approx. 200,000 digital process data sets result per active operating hour. Meanwhile (as at: December 2014), approx. 4 million process data sets are saved. Irrespective of the complexity of the analysis and the number of data sets, only about 3 s is always required for one query. Google achieves this by corresponding server clustering, i.e. more and more server are used for the evaluation as the data volume rises. Viewed overall, however, the data and analysis volume is very low in relation to the potential of Google BigQuery, with the result that no costs were accrued for storage and analysis up to now.

The following two examples show the current analyses at the demonstration system, which were used for optimizing the work on the system.

The analysis according to Figure 7 shows the time of day for the highest productivity, i.e. at which time of the day the most model cars are assembled at the system (number of transporters).

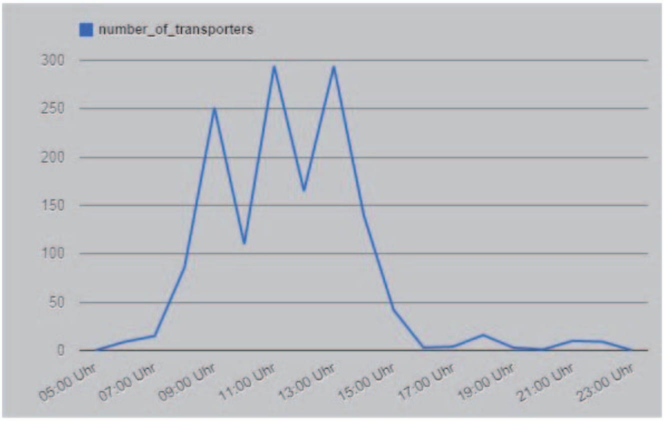

Fig. 7. Time of day for the highest productivity at the demonstration system

The gripper movement at stations 2, 4 and 5 was examined in Figure 8.

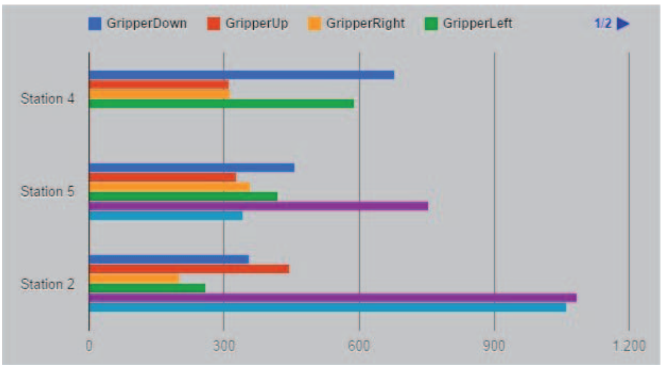

Fig. 8. Movement times of the grippers at stations 2, 4 and 5

The movement times of the gripper arms at the stations are between (200 ... 1100) ms. Even if the storage of the individual process values via the IP network have a certain scatter [10], correct values for the movement times are determined by analyzing a large volume of data in the statistical average. Based on the measured values, the movement times of the gripper arms were reset via pressure reducing valves for the three stations. As a result, the gripper movement times at the assembly stations could be optimized and an increase in the productivity and a wear reduction of the actuation valves could be attained. Without using the BD services, a recalibration of the valves at the individual stations would have meant an extensive stationary measurement setup with involved time measurements. A permanent tracking of the conditions in the stations would also not be possible for a preventative diagnosis/maintenance.

5. FURTHER WORKS

Further data analyses are intended at the demonstrator for the winter semester 2014/15. About 50 students will carry out project work at the assembly stations of the demonstrator during the semester (development of browser-based operator panel for the stations by use of remote access and Web collaboration). The BD store service will record the process signals of the stations during the entire semester, with the result that a performance analysis can be conducted for the individual stations.

A new R&D project "Potential, Structure and Interfaces of Big Data Services for CPS-based Automation" is also planned for 2015. The project is to examine the options and application potential of Big Data analysis for RTP data from automated systems and research a suitable reference architecture based on the service paradigm for this. The project results are to reveal the potentials and risks of Big Data analysis for very large amounts of RTP data ($> 10^{10}$ data sets) as well as the required structures and interfaces along with possible realization variants.

REFERENCES

[1] Kagermann, H., Wahlster, W., Helbig, J.: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Abschlussbericht des Arbeitskreises Industrie 4.0 Forschungsunion Wirtschaft – Wissenschaft, 2013

[2] Kuppinger, S.: Cloud in der Automation. http://www.computer-automation.de/steuerungsebene/industrie-c/news/article/86246/0/Cloud_in_der_Automation/, 2013

[3] Ribaric, A.: Embracing the Future of the Industrial Internet – Prognostics and the Next Wave of Data Analytics. – Whitepaper, Sentient Science, New York, 2014

[4] CCAD: Final project report WOAS. http://195.37.242.50/downloads/WOAS/WOAS_Schlussbericht_2014.pdf, 2014

[5] Langmann, R.: Web-Connector für OPC. In: A&D Kompendium 2014/2015, pp. 84 – 87. publish-industry Verlag, Munich, 2013

[6] BSI: Cloud Computing Grundlagen. - https://www.bsi.bund.de/DE/Themen/CloudComputing/Grundlagen/Grundlagen_node.html, 2014

[7] Brass, M.: Anwendung von Google Analytics zur Auswertung von Prozessdaten. Documentation of the Semester project, Duesseldorf University of Applied Sciences, CCAD, 2014

[8] Langmann, R. and Meyer, L.: Architecture of a Web-oriented Automation System. – 18th IEEE Int. Conf. on Emerging Technologies & Factory Automation (ETFA 2013), 10. – 13.09.13, Cagliari, Italy, Proceedings

[9] CCAD: WOAD-Interfacebeschreibung. Developer documentation, Duesseldorf University of Applied Sciences, CCAD, 2014

[10] Langmann, R.: Ein Interface für CPS-basierte Automatisierungsgeräte. In: Proceedings of Konferenz für Angewandte Automatisierungstechnik in Lehre und Entwicklung (AALE 2014), pp. 133 – 142. DIV Deutscher Industrieverlag, Munich, 2014