

Big Data-Driven Industry 4.0 Service Engineering Large-Scale Trials: The Boost 4.0 Experience



Oscar Lázaro, Jesús Alonso, Paulo Figueiras, Ruben Costa, Diogo Graça, Gisela García, Alessandro Canepa, Caterina Calefato, Marco Vallini, Fabiana Fournier, Nathan Hazout, Inna Skarbovsky, Athanasios Poulakidas, and Konstantinos Sipsas

Abstract In the last few years, the potential impact of big data on the manufacturing industry has received enormous attention. This chapter details two large-scale trials that have been implemented in the context of the lighthouse project Boost 4.0. The chapter introduces the Boost 4.0 Reference Model, which adapts the more generic BDVA big data reference architectures to the needs of Industry 4.0. The Boost 4.0 reference model includes a reference architecture for the design and implementation of advanced big data pipelines and the digital factory service development reference architecture. The engineering and management of business network track and trace processes in high-end textile supply are explored with a

O. Lázaro (✉) · J. Alonso

Asociación de Empresas Tecnológicas Innovalia, Derio, Spain
e-mail: olazaro@innovalia.org; jalonso@innovalia.org

P. Figueiras · R. Costa

UNINOVA- Instituto de desenvolvimento de novas tecnologias – Associacao, Caparica, Portugal
e-mail: paf@uninova.pt; rddc@uninova.pt

D. Graça · G. Garcia

Volkswagen Autoeuropa, Anjo, Portugal
e-mail: diogo.graca@volkswagen.pt; gisela.garcia@volkswagen.pt

A. Canepa

Fratelli Piacenza, Pollone, Italy
e-mail: alessandro.canepa@piacenza1733.it

C. Calefato · M. Vallini

Domina Srl, Ragusa, Italy
e-mail: caterina.calefato@domina-biella.it; marco.vallini@domina-biella.it

F. Fournier · N. Hazout · I. Skarbovsky

IBM Israel, Haifa, Israel
e-mail: fabiana@il.ibm.com; nathanh@il.ibm.com; inna@il.ibm.com

A. Poulakidas · K. Sipsas

Intrasoft International S.A., Luxembourg City, Luxembourg
e-mail: Athanasios.Poulakidas@intrasoft-intl.com; Konstantinos.SIPSAS@intrasoft-intl.com

© The Author(s) 2022

E. Curry et al. (eds.), *Technologies and Applications for Big Data Value*,
https://doi.org/10.1007/978-3-030-78307-5_17

focus on the assurance of Preferential Certification of Origin (PCO). Finally, the main findings from these two large-scale piloting activities in the area of service engineering are discussed.

Keywords Reference architecture · ISO 20547 · ISO/IEC/IEEE 42010 · DIN 27070 · Sovereignty · Data spaces · Track & Trace · Blockchain · FIWARE · Virtual commissioning · Testbed · Trial · Business networks 4.0 · SUMA 4.0 · Intralogistics

1 Introduction

Over the last few years, the potential impact of big data for the manufacturing industry has received enormous attention. However, although big data has become a trend in the context of manufacturing evolution, there is not yet sufficient evidence on how and if big data will leverage such impact in practical terms. New concepts in the area of Industry 4.0 such as digital twins, digital threads, augmented decision support dashboards and systems, and simulation-based commissioning systems rely significantly on advanced engineering and operation of big data techniques and technical enablers. The emergence of data-driven techniques to increase data visibility, analytics, prediction and autonomy has been immense. However, those techniques have been developed in many cases as individual efforts, without the availability of an overarching framework making the transfer of such applications to other industries at scale cumbersome. Moreover, the development of such big data applications is not necessarily realized in context with reference architectures such as the European Reference Architectural Model Industry 4.0 (RAMI 4.0), which serves as reference in the sector for Industry 4.0 digital transformation. Big data promises to impact Industry 4.0 processes at all stages of the product life-cycle.

The aim of this chapter is to present the advances made in the area of service engineering and commissioning in the context of H2020 EU large-scale piloting project Boost 4.0 [1]. It gathers the first set of experiences, best practices and lessons learned during the deployment of the two lighthouse trials in the scope of the Boost 4.0 project: the most ambitious European initiative in big data for Industry 4.0. It presents the experiences of two European manufacturing leaders (large industry and SME) in the engineering and management at large scale of data-driven and traceable intra-logistics and supply chain processes. Intra-logistic processes will be addressed by the Volkswagen Autoeuropa (Portugal) plant in the automotive sector, whereas supply chain business network engineering and management will be addressed by the Italian SME Piacenza in the high-end textile sector. This chapter addresses initial data value innovation elicitation and presents and assesses how a common RA can be used to leverage advanced service engineering practices at large scale, as well as lessons learned and impact evaluation.

This chapter relates mainly to the technical priorities Data Management Engineering and optimized architectures for analytics of data-at-rest and data-in-motion

of the European Big Data Value Strategic Research & Innovation Agenda [2]. It addresses the horizontal concerns of heterogeneity, scalability and processing of data-in-motion and data-at-rest of the BDV Technical Reference Model. It addresses the vertical concerns of communication and connectivity, engineering and DevOps for building big data value systems areas to facilitate timely access and processing of big data and evolving digital twin models. The work in this chapter relates mainly but not only to the Systems, Methodologies, Hardware and Tools cross-sectorial technology enablers of the AI, Data and Robotics Strategic Research, Innovation & Deployment Agenda [3].

The chapter is organized as follows: First the Boost 4.0 initiative is introduced with a focus on the instantiation of the Boost 4.0 common big data-driven Reference Architecture (RA). This RA is aligned with the big data RA proposed by Big Data Value Association (BDVA) and harmonized with the Digital Factory Alliance (DFA) overall digital factory open reference model. Next, the big data intra-logistic process engineering trial and lessons learned at Volkswagen Autoeuropa are introduced. Next, the engineering and management of business network track and trace processes in high-end textile supply are presented with a focus on assurance of Preferential Certification of Origin (PCO). Finally, the main findings extracted from these two large-scale piloting activities in the area of service engineering are discussed.

2 Boost 4.0 Universal Big Data Reference Model

Boost 4.0 (Big Data Value Spaces for Competitiveness of European Connected Smart Factories 4.0) is the biggest European initiative in big data for Industry 4.0. With a 20 M€ budget and leveraging 100 M€ of private investment, Boost 4.0 has led the construction of the European Industrial Data Space to improve the competitiveness of Industry 4.0. Since January 2018, it has guided the European manufacturing industry in the introduction of big data in the factory, providing the industrial sector with the necessary tools to obtain the maximum benefit of big data.

Since the beginning of the project, Boost 4.0 has demonstrated in a realistic, measurable and replicable way an open, certifiable and highly standardized shared data-driven Factory 4.0 model through 11 lighthouse factories, and has also demonstrated how European industry can build unique strategies and competitive advantages through big data across all the phases of product and process lifecycle.

2.1 Boost 4.0 Objectives

Boost 4.0's overall mission is to accelerate the adoption of Industry 4.0 big data-intensive smart manufacturing services through highly replicable lighthouse activities that are intimately connected to current and future Industry 4.0 invest-

ments, resolving the smart connected product and process data fragmentation and leveraging the Factory 4.0 data value chain.

To accomplish this mission, Boost 4.0 has defined the following objectives:

- **Global Standards:** Contribution to the International Data Space data models and open interfaces aligned with the European Reference Architectural Model Industry 4.0 (RAMI 4.0).
- **Secure Digital Infrastructure:** Adaptation and extension of cloud and edge digital infrastructures to ensure high-performance operation of the European Industrial Data Spaces, i.e. support of high-speed processing and analysis of huge and very heterogeneous industrial data sources.
- **Trusted Big Data Middleware:** Integration of the four main open-source European initiatives (International Data Space, FIWARE, Hyperledger, Big Data Europe) to support the development of open connectors and big data middleware.
- **Digital Manufacturing Platforms:** Opening of interfaces for the development of big data pipelines for advanced analysis services and data visualization supported by the main digital engineering, simulation, operations and industrial quality control platforms.
- **Certification:** Development of a European certification programme for equipment, infrastructures, platforms and big data services for operation in the European Industrial Data Space.

2.2 *Boost 4.0 Lighthouse Factories and Large-Scale Trials*

In Boost 4.0, some of the most competitive factories, from three strategic economic sectors that drive not only European manufacturing economy but also the IoT/smart connected market development (i.e. automotive, manufacturing automation and smart home appliance sectors) join forces to set up 11 lighthouse factories and 2 replication factories (Fig. 1) that are a coherent, complementary and coordinated big data response to the 5 EFFRA Factory 4.0 Challenges, i.e. (1) lot size one distributed manufacturing, (2) operation of sustainable zero-defect processes and products, (3) zero break down operations, (4) agile customer-driven manufacturing value network management and (5) human-centred manufacturing.

Boost 4.0 leverages five widely applicable big data transformations: (1) networked commissioning and engineering, (2) cognitive production planning, (3) autonomous production automation, (4) collaborative manufacturing networks and (5) full equipment and product availability—across each of the five key product and process lifecycle domains considered: (1) Smart Digital Engineering, (2) Smart Production Planning and Management, (3) Smart Operations and Digital Workplace, (4) Smart Connected Production and (5) Smart Maintenance and Service.

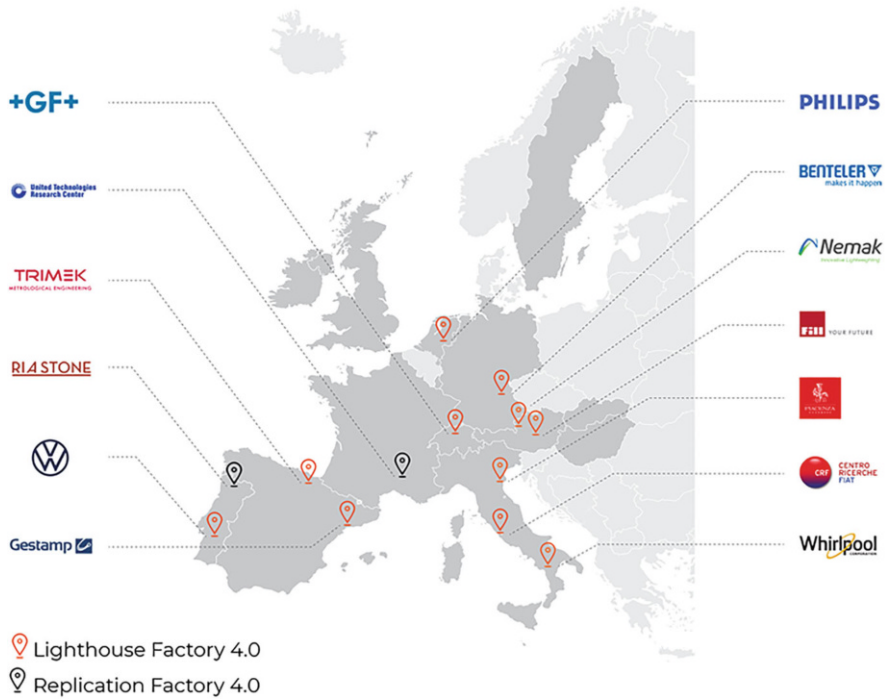


Fig. 1 BOOST 4.0 big data driven lighthouse and replication factories 4.0

2.3 Boost 4.0 Universal Big Data Reference Architecture

One of the main ambitions of Boost 4.0 is to define and develop highly replicable big data solutions to ensure the impact of the project beyond the project lifetime. One of the main challenges Industry 4.0 faces when designing their big data solutions is first of all to effectively address the design and development of high-performance big data pipelines for advanced data visualization, analytics, prediction or prescription. Then, the challenge lies in how to successfully integrate such big data pipelines in the digital factory engineering and production frameworks. In this sense, to facilitate the replicability of the lighthouse trials and big data solutions implemented, Boost 4.0 has relied on two reference models. On one hand, the BDVA Big Data Reference Model (BD-RM) [4] to drive Industry 4.0 big data pipelines and process engineering and operation. The goal of this RM is to ensure universality and transferability of trial results and big data technologies as well as economies of scales for big data platform and technology providers across sectors.

On the other hand, Boost 4.0 has developed and applied a RAMI 4.0 [5] compliant Service Development Reference Architecture (SD-RA) for big data-driven factory 4.0 digital transformation. This model is now maintained by the Digital Factory Alliance (DFA) [6]. The goal is to ensure a perfect alignment

between big data processes, platforms and technologies with overall digital transformation and intelligent automation efforts in manufacturing factories and connected manufacturing networks.

As illustrated in Fig. 2, the Boost 4.0 BD-RA [7] is composed of four main layers: Integration Layer, Information and Core Big Data Layer, Application and Business Layers. This approach is aligned with the ISO 20547 Big Data Reference Architecture—Big Data Application Provider layer—from Data Acquisition/Collection through Data Storage/Preparation (and sharing) further to any Analytics/AI/Machine Learning and also environmental Action/Interaction including Visualization.

These four layers allow the implementation of a big data pipeline and the integration of such pipelines in specific business processes supporting the Factory 4.0 product, process and service lifecycle, i.e. smart digital engineering, smart digital planning and commissioning, smart digital workplace and operations, smart connected production and smart servicing and maintenance. These four Boost 4.0 layers are supported by a set of transversal services, in particular data sharing platforms, engineering and DevOps, Communications and Networking, Standards and Cybersecurity and Trust. These layers enact the manufacturing 4.0 entities and leverage a data 4.0 value chain that transforms raw data sources into quality data that can be interpreted and visualized, providing mining and context for decision support. This value chain is developed as data is aggregated, integrated, processed, analysed and visualized across the Factory 4.0 layers (product, device, station, workcentre, enterprise and connected world). The Boost 4.0 BD-RA adopts the BDVA RM and adapts it to the specific needs of Industry 4.0.

However, the generic Boost 4.0 BD-RA needs to be articulated and instantiated with the support of specific platforms, solutions and infrastructures so that the big data-driven manufacturing processes can actually be realized. So, even if, as shown in Fig. 2, the BDVA big data reference model can in fact be adapted to Industry 4.0 needs and aligned with the RAMI 4.0 model, a more formal harmonization and integration of the BDVA RM is required to facilitate development of big data services in the context of a digital factory exhibiting high transferability and replication capabilities for big data-driven manufacturing processes. This is further facilitated with the application of the DFA Digital Factory Service Development RA (SD-RA), which ensures a broad industrial applicability of digital enablers, mapping the technologies to different areas and to guide technology interoperability, federation and standard adoption. The DFA SD-RA design complies with ISO/IEC/IEEE 42010 [7] architectural design principles and provides an integrated yet manageable view of digital factory services. In fact, DFA SD-RA integrates functional, information, networking and system deployment views under one unified framework. The DFA SD-RA address the need for an integrated approach to how (autonomous) services can be engineered, deployed and operated/optimized in the context of the digital factory. With this aim, the DFA SD-RA is composed of three main pillars, as depicted in Fig. 3:

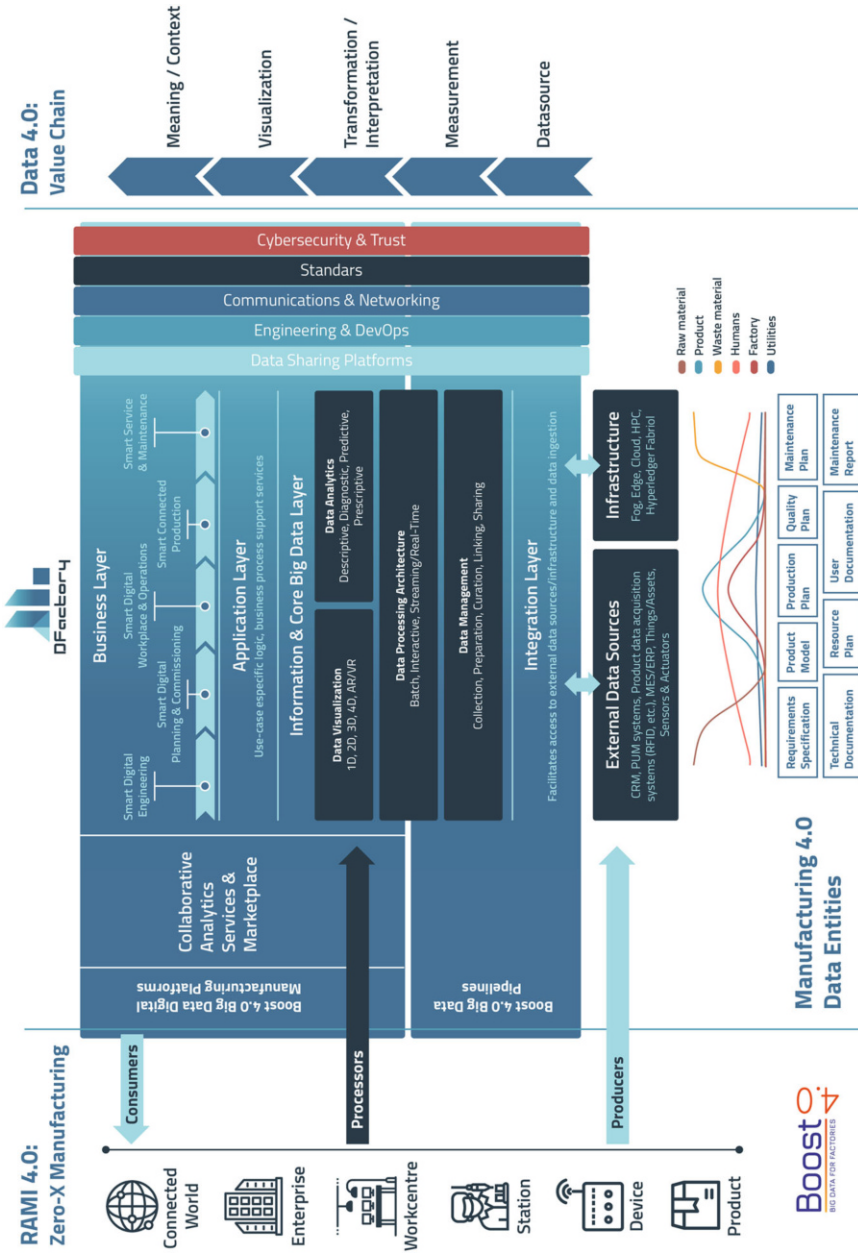


Fig. 2 Boost 4.0 Universal Big Data Reference Architecture 4.0

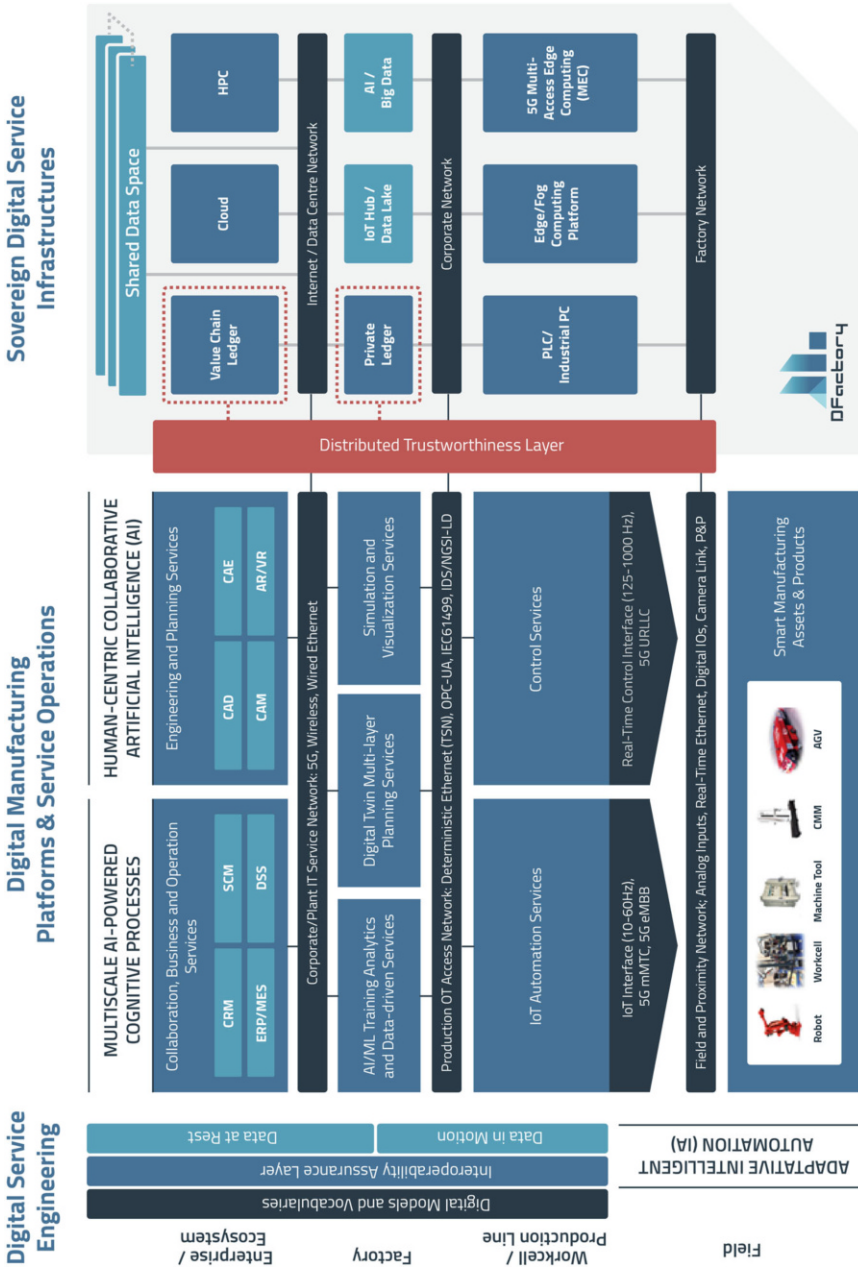


Fig. 3 Digital Factory Alliance Reference Architecture for Industry 4.0

1. **Digital Service Engineering.** This pillar provides the capability in the architecture to support *collaborative model-based service enterprise approaches* to digital service engineering of (autonomous) data-driven processes with a focus on supporting smart digital engineering and smart digital planning and commissioning solutions to the digital factory. The pillar is mainly concerned with the harmonization of digital models and vocabularies. It is this pillar that should develop interoperability assurance layer capabilities with a focus on mature digital factory standards adoption and evolution towards an “industry commons” approach for acceleration of big data integration, processing and management. It is this pillar where “security by design” can be applied both at the big data, manufacturing process and shared data space levels.
2. **Digital Manufacturing Platforms and Service Operations.** This pillar supports the deployment of services and DMPs across the different layers of the digital factory to enact data-driven smart digital workplaces, smart connected production and smart service and maintenance manufacturing processes. The pillar is fundamental in the development of three enabling capabilities central to the gradual evolution of autonomy in advanced manufacturing processes, i.e. multi-scale AI-powered cognitive processes, human-centric collaborative intelligence and adaptive Intelligent Automation (IA). The enablement of both knowledge-based (multi-scale artificial intelligence) and data-driven approaches (collaborative intelligence) to digital factory intelligence is facilitated by the support of service-oriented and event-driven architectures (interconnected OT and IT interworking event and data buses) embracing international and common standard data models and open APIs, thereby enabling enhanced automated context development and management for advanced data-driven decision support.
3. **Sovereign Digital Service Infrastructures.** The operation of advanced digital engineering and digital manufacturing platforms relies on the availability of suitable digital infrastructures and the ability to effectively develop a *digital thread* within and across the digital factory value chain. DFA SD-RA relies on infrastructure federation and sovereignty as the main design principles for the development of the data-driven architecture. This pillar is responsible for capturing the different digital computing infrastructures that need to be resiliently networked and orchestrated to support the development of different levels and types of intelligence across the digital factory. In particular, the DFA SD-RA considers three main networking domains for big data service operation; i.e. factory, corporate and internet domain. Each of these domains needs to be equipped with a suitable security and safety level so that a seamless and cross-domain distributed and trustworthy computing continuum can be realized. The pillar considers from factory-level digital infrastructure deployment such as PLC, industrial PC or Fog/Edge to the deployment of telecom-managed infrastructure such as 5G multi-access edge computing platforms (MEP). At the corporate level, the reference architecture addresses the need for the development of IoT Hubs that are able to process continuous data streams as well as dedicated big data lake infrastructures, where batch processing and advanced analytic/learning services can be implemented. It is at this corporate level that private ledger

infrastructures are unveiled. Finally, at the internet or data centre level, the digital factory deploys advanced computing infrastructures exploiting HPC, Cloud or value chain ledger infrastructures that interact with the federated and shared data spaces.

The DFA RA is aligned with ISO 20547 Big Data Reference Architecture. The DFA Sovereign Digital Service Infrastructures pillar allows Boost 4.0 reference model to additionally address the ISO 20547 Big Data Framework Provider layer. The DFA RA is composed of four layers that address the implementation of the 6 big data “C” (Connection, Cloud/edge, Cyber, Context, Community, Customization), enables four different types of intelligence (smart asset functioning, reactive reasoning, deliberative reasoning and collaborative decision support) to be orchestrated and maps to the 6 layers of the RAMI 4.0 (product, devices, station, workcentre, enterprise and connected world), which target all relevant layers required for the implementation of AI-powered data-driven digital manufacturing processes:

1. The lower layer of the DFA RA contains the *field devices in the shopfloor*: machines, robots, conveyer belts as well as controllers, sensors and actuators are positioned. Also in this layer the smart product would be placed. This layer is responsible for supporting the development of different levels of autonomy and *smart product and device (asset) services* leveraging on intelligent automation and self-adaptive manufacturing asset capabilities.
2. The *workcell/production line* layer represents the individual production line or cell within a factory, which includes individual machines, robots, etc. It covers both the services, that can be grouped in two those that provide information about the process and the conditions (*IoT automation services*), and the actuation and control services (*automation control services*); and the infrastructure, typically represented in the form of PLC, industrial PCs, edge and fog computing systems or managed telecom infrastructures such as MEC. This layer is responsible for developing reactive (fast) reasoning capabilities (automated decision) in the SD-RA and leveraging augmented distributed intelligence capacities based on enhanced management of context and cyber-physical production collaboration.
3. At the *factory level*, a single factory is depicted, including all the work cells or production lines available for the complete production, as well as the factory-specific infrastructure. Three kinds of services are typically mapped in this layer: (1) *AI/ML training, analytics and data-driven services*; (2) *digital twin multi-layer planning services*; and (3) *simulation and visualization services*. The infrastructure that corresponds to this layer is the IoT Hubs, data lakes and AI and big data infrastructure. This layer is responsible for supporting the implementation of deliberative reasoning approaches in the digital factory with planning (analytical, predictive and prescriptive capabilities) and orchestration capabilities, which combine and optimize the use of analytical models (knowledge and physics based), machine learning (data-driven), high-fidelity simulation (complex physical model) and hybrid analytics (combining data-driven and model-based methods) under a unified computing framework. This

leverages in the architecture collaborative assisted intelligence for explainable AI-driven decision processes in the manufacturing environment.

4. The higher layer refers to the enterprise/ecosystem level, that encompasses all *enterprise and ecosystem* (connected world) services, platforms and infrastructures as well as interaction with third parties (value chains) and other factories. The global software systems that are common to all the factories (*collaboration business and operation services* as well as *engineering and planning services*) are supported usually by Cloud or HPC infrastructures. It is this layer that supports the implementation of shared data spaces and value-chain-level distributed ledger infrastructures for implementation of trusted information exchange and federated processing across *shared digital twins* and asset administration shells (AAS). This layer leverages a human-centric augmented visualization and interaction capability in the context of data-driven advanced decision support or generative manufacturing process engineering.

2.4 Mapping Boost 4.0 Large-Scale Trials to the Digital Factory Alliance (DFA) Service Development Reference Architecture (SD-RA)

This chapter aims to present two Boost 4.0 lighthouse trials that focus on the engineering and process planification services, using big data technologies and exploiting the digital twin capabilities to improve the overall production process (Fig. 4). Each section corresponds to one trial:

Section 2 describes the trial that was deployed in Volkswagen Autoeuropa Plant in Palmela (Portugal). This lighthouse factory has deployed a big-data-based solution to plan intra-logistic processes, which fully integrates the material flow from the unloading docks to the point of fit.

Section 3 introduces the Piacenza lighthouse trial, discussing how a business network can be developed in the high-end textile sector with the support of blockchain technology to guarantee traceability and visibility through the supply chain.

3 Big Data-Driven Intra-Logistics 4.0 Process Planning Powered by Simulation in Automotive: Volkswagen Autoeuropa Trial

Volkswagen Autoeuropa (VWAE) belongs to an automotive manufacturing industry located in Portugal (Palmela) since 1995 and is a production plant of Volkswagen Group. VWAE plays a strategic role in the Portuguese automotive industry, as it is the largest automotive manufacturing facility in the country and is responsible for

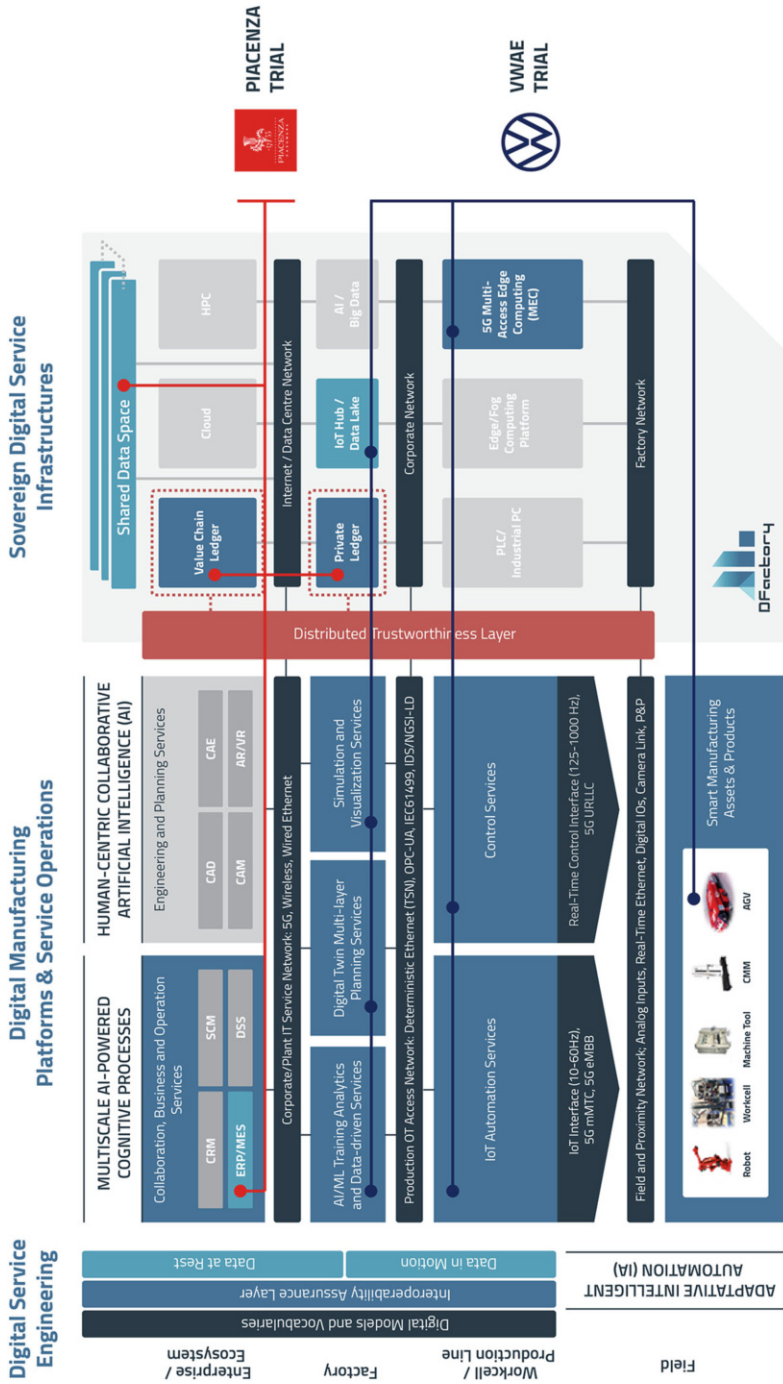


Fig. 4 Mapping of the two service engineering trials to the DFA SD-RA

around 10% of all Portuguese exportations. The plant employs around 6000 workers and, indirectly, it employs close to 8000 people through the more than 800 suppliers that provide materials, components and parts to the facility.

The goal of VWAE, within the Boost 4.0 project, is to take advantage of the latest big data technology developments and apply them to an industry environment with non-stop cycles and with high up-times. In the end, the desirable target is to transform an environment overwhelmed with manual complex processes with one that brings modular flexibility and automation.

The expected benefits with the implementation of a data-driven autonomous warehouse would translate into financial benefits for the Volkswagen Group, increase in flexibility (which is key specially during the introducing of a new model), minimization of human dependency for manual operations and, thus, an increase in the process efficiency. The automation and control of the process through a big data architecture enables a business intelligence approach to the warehouse system. Tools such as reporting, Digital Twin simulation, monitoring and optimization-support offer the opportunity to analyse and improve the system with real-world big data.

3.1 Big Data-Driven Intra-Logistic Planning and Commissioning 4.0 Process Challenges

Currently the logistics process is heavily reliable on manual processes and in addition to that, the operation is performed inside the factory, where space is limited. On the receiving area, trucks are traditionally unloaded by a manual forklift operation, and then the unit loads are transported to the warehouse where they will be stored either in shelves or block storage concept. System wise there is one database to control the parts coming from each truck and then a separate database which registers the unloading, transportation and storing of the material in the warehouse.

Figure 5 represents the data silos used throughout the process to collect the necessary logistics information. Besides the labour-intensive tasks within the logistics at VWAE, the data silo-based architecture does not allow the monitorization and optimization of the overall logistics process. Apart from the data silos for receiving, unloading, warehousing and sequencing, there is a lack of information about the transport operations between these phases in the process. Furthermore, data is captured and collected manually, which contributes to loss of time in the process and potentiates the existence of errors in the collected data.

Hence, the main challenge is to transform the siloed nature of data storage within the logistics process to support a true big data architecture, from which valuable insights can be extracted so as to optimize the whole logistics process and to aid in the optimization and automation efforts within the logistics process at VWAE. To achieve the transformation to a big data context, the integration of data present in the various silos is of the utmost importance. Such data integration efforts will enable the application of big data processing and analytics methods that will support

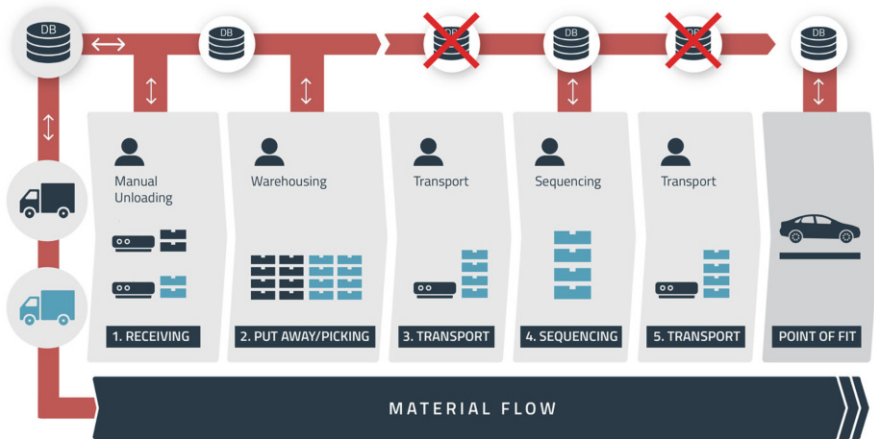


Fig. 5 Intra-logistic silo-based system flow of current process

the capitalization on valuable insights within the process. Moreover, the envisaged big data architecture will also form a basis for the development of a digital twin of the logistics area, which will enable real-world simulation, testing and validation of new automated solutions without the need for actual application in real-world, ready-for-production scenarios.

The planning and commissioning of advanced intra-logistics 4.0 processes therefore presents clear big data challenges in the velocity (real-time warehouse data streaming), veracity (accuracy of digital twin simulations), variety (breaking intralogistics information silos) and volume (data deluge) dimensions.

3.2 Big Data Intra-Logistic Planning and Commissioning Process Value

The expected future scenario aims at achieving a full integration of the material flow, from receiving up to the point of fit. Figure 6 shows the system flow integration as it is foreseen in VWAE. The main objective of the VWAE trial is to eliminate human intervention or at least reduce to a minimum at all phases from receiving up to the point of fit.

In order to test and validate the future scenario, a recurrent issue was chosen as a proof-of-concept: the issue of optimum stock in the logistics area. Due to the lack of data-supported, informed decisions in the process of supply ordering, the logistics area is often in a situation of overstock, meaning that there is always a surplus of parts that goes beyond the envisaged safety stock. The safety stock exists to tackle problems of parts' delivery, due to transportation issues or other obstacles, such as supplier shortfalls due to demand instability. Overstock has several consequences, from overspending and time-in-shelf issues to more concrete problems, such as part rejection due to its temporal validity.

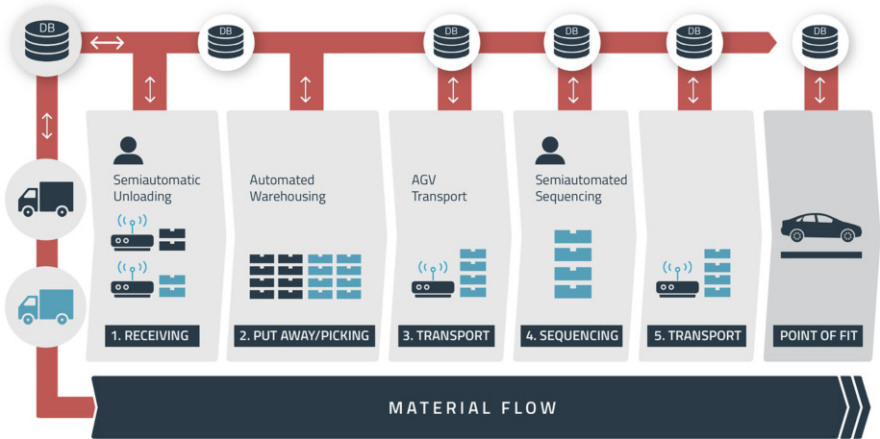


Fig. 6 System flow of data-driven intra-logistic 4.0 process

Hence, the chosen proof-of-concept was the overstock of batteries, since batteries are perishable parts (they have temporal validity) and the overstock situations for this type of part is a known problem and mitigating it represents real business value due to the unit price involved.

3.3 *Big-Data Pipelines for Intra-Logistic Planning and Commissioning Solutions in Automotive*

Figure 7 shows the general big data architecture and core open source big data technologies that support most of data ingestion, processing and management work, namely to efficiently gather, harmonize, store and apply analytic techniques to data generated within the intra-logistics process. The use of big data technologies with parallel and distributed capabilities is essential to address the processing of large batch/stream data with different levels of velocity, variety and veracity. Therefore, the architecture must meet requirements such as scalability, reliability and adaptability.

The architecture is mainly split into four layers: Data ingestion layer, Data Storage layer, Data Processing layer, and Data Querying/Analytics/Visualization layer. For data processing and collection, Apache Spark [8] is used in conjunction with the IDSA Connectors [9], enabling direct linkage with the IDSA Ecosystem, while for big data storage, the chosen technologies were PostgreSQL [10] and MongoDB [11]. Finally, for data querying and access, data analytics and data visualization, the chosen tools were, respectively, Apache Hive [12], Spark Machine Learning Library and Grafana [13].

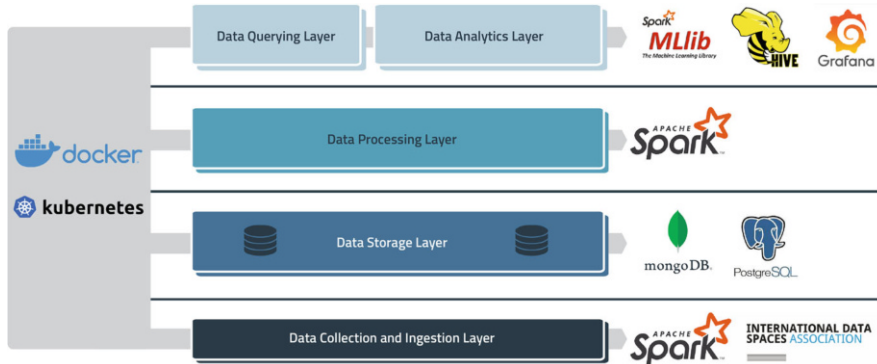


Fig. 7 Big data architecture for the VWAE trial

3.4 Large-Scale Trial of Big Data-Driven Intra-Logistic Planning and Commissioning Solutions for Automotive

The large-scale trial connects the Visual Components simulation environment with the suite of big data Analytics and Machine Learning tools, provided by UNINOVA, in a bidirectional way, as shown in Fig. 8: First, big data and machine learning technologies are used to aggregate real-time logistics operations' data, perform prediction over the data if needed, and send the results to the Visual Components simulation environment. Then, after the simulation ends, analytics and machine learning techniques are used in order to analyse key performance indicator data returned by the simulation environment, in order to find patterns, anomalies or possible points of optimization for future reference.

The 3D simulation environment replicates the trial scenario within the virtual world, i.e. a digital twin. The real model provides the logistics process data, which, after simulation, are validated with the current production outcomes. Once the simulation scenario is validated, simulation data can be analysed to be reused in the simulation to improve process performance and building the digital twin.

When the first version of the simulation, or digital twin, was developed, there was a need to propose actual key performance indicators (KPIs) extracted from the real logistics processes, focusing on the arrival and storage of batteries, with the simulation itself. Several KPIs were selected, such as the number of batteries, per battery type, in the warehouse and in the sequencing area at any given time, the occupation percentage of workers in the several logistics steps and the overall execution times of the different processes.

The first KPI to be validated in this phase was the reduction of truck arrivals, and consequent decrease of battery palettes in stock. The reduction in KPI corresponds to a decrease of 5% of the stock for the so-called high runners: the types of batteries that are most used in the production line. So, the test was performed as follows:

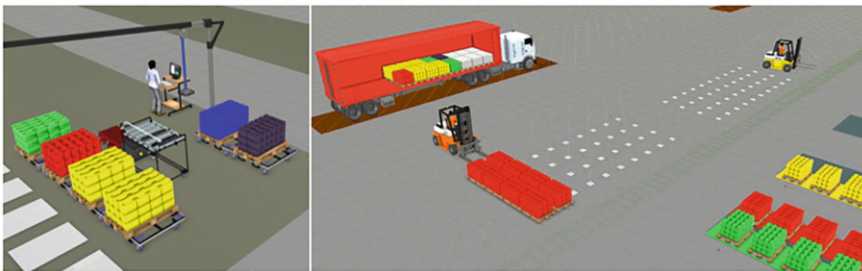
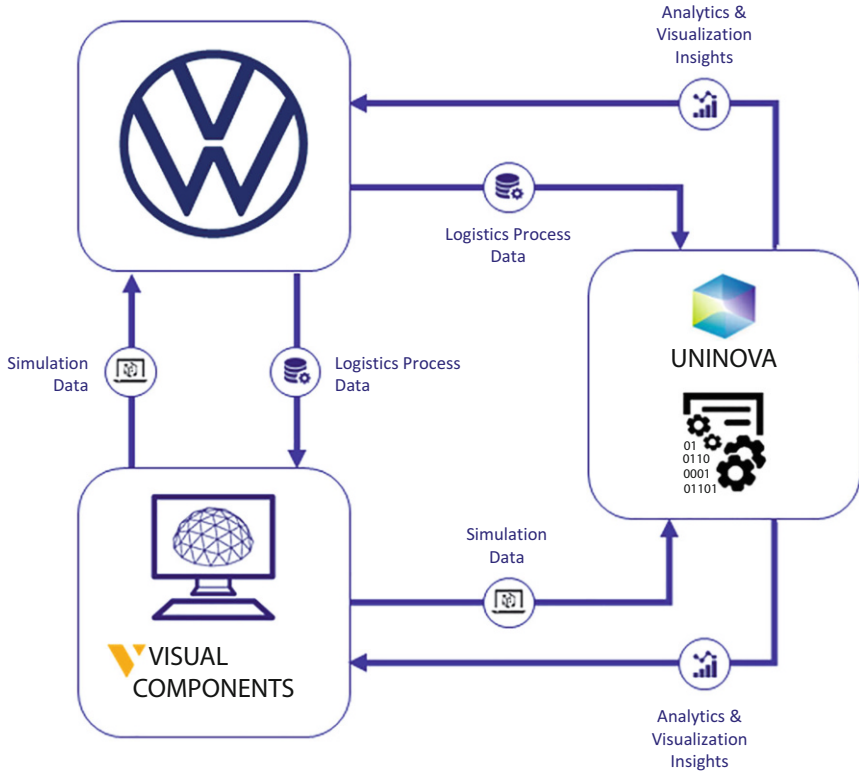


Fig. 8 VWAE Digital Twin Analytics trial data flow and planning platform

1. Real data corresponding to the truck arrivals, and to the usage of batteries in production was injected into the simulation, via Orion Context Broker. From this data injection, the Digital Twin produced a benchmark for the battery palettes' arrival percentages and truck arrival times.
2. The selected KPI was to decrease the arrival of high-runner palettes by 5%, while increasing the time interval between trucks, also saving in CO₂ emissions and direct costs for transport and stock, but maintaining the current production rates. The percentages of arriving palettes were arranged so that there would be a cut of

5% in the high runners while maintaining a total throughput/arrival of 100%. The time between truck arrivals was also arranged in order to have bigger intervals.

3. Hence, an average decrease of 5% in the high runners' arrival percentage, along with an increase in the truck arrival time interval was simulated. The new values showed that the battery stock was always above the level required, even with the decrease of 5% arrival of batteries and the increase in the time interval between truck arrivals.
4. Finally, a prediction model for the battery stock optimization was developed and tested. The chosen model was an optimized long short-term memory (LSTM) which is an artificial recurrent neural network model. This choice was made because LSTM are reportedly very good at forecasting time series data and do not require a lot of parameterization for multivariate datasets. Historical data was used to estimate the possible optimizations.

In 2018, there were multiple cases of overstock of car batteries at VWAE. For instance, in the case of the batteries, the warehouse was at least half of the time in overstock situations and 25% of the time in severe overstock. The results showed that a significant decrease in stock can be achieved, along with real benefits for VWAE, financially, by cutting in stock costs, and environmentally, by reducing both the number of truck arrivals and the occurrence of past-validity batteries.

3.5 Observations and Lessons Learned

The fusion between the big data architecture, developed in the Boost 4.0 project, and the Visual Components simulation environment, in order to create a true Digital Twin, was proven to be a crucial decision-support system, in the sense that it helped relevant stakeholders at VWAE to better understand the limitations in the current logistics process, but also to optimize critical aspects of this process, such as in the case of the overstock situation.

Furthermore, the overall system is ready for full scale-up, since it is capable of ingesting data from the whole logistics process, and for all the parts that are necessary for automotive production. The system is also ready to simulate, in near-real-world conditions, all phases of the logistics process, apart from the arrival of trucks. Hence, it will be a powerful aid in achieving the future automation requisites of VWAE, by enabling the simulation of new, automated and optimized scenarios for the logistics processes.

4 From Sheep to Shop Supply Chain Track and Trace in High-End Textile Sector: Piacenza Business Network Trial

Piacenza company is based in the Italian textile district of Biella, where all its production is carried out, and is one of the oldest textile industries in the world, founded in 1733 and from then on owned by the Piacenza family. Piacenza is one of the few undisputed worldwide leaders in high fashion fabrics and accessories production, with a competitive strategy focused on the maximum differentiation of the product, in terms of raw material choice, style, and colour. Fabric production includes more than 70 production passages or steps, which starts in the countries of origin of the natural fibres used for fashion fabrics (cashmere, vicuna, alpaca, mohair, silk, wool, linen, etc.) and can be summarized into three main changes of material status: raw material → yarn → fabric.

High textile fabric production is characterized by an extremely high number of product variables, deep customization, hardly predictable demand, length of production cycle (60–90 days from raw materials to receipt), physical prototyping and sampling, fragmented distribution and very small batches due to high customization. The combination of these aspects leads to a very complex production, which must properly balance the request of a very fast and demanding market with the length and rigidity of a fragmented and long value chain.

4.1 Data-Driven Textile Business Network Tracking and Tracing Challenges

The garment and footwear industry has one of the highest environmental footprints and risks for human health and society. The complexity and opacity of the value chain makes it difficult to identify where such impacts occur and to devise necessary targeted actions. Key actors in the industry have identified interoperable and scalable **traceability** and **transparency** of the value chain, as crucial enablers of more responsible production and consumption patterns, in support of Sustainable Development.

—United Nations Economic and Social Council [14].

Textile and clothing play a significant role in climate change with 1.7 million tons/year of CO₂ emissions [15], 10% of substances of potential concern to human health, 87% of the workforce (mainly women) gets below living wages. Permitted by lowered cost, a garment is worn an average of 3 times in its lifecycle, with 400 billion euros lost a year due to discarding clothes which can still be worn, 92 million tons of fashion waste every year, 87% of clothes ending up in landfills. In addition, the market for counterfeit clothing, textiles, footwear, handbags, cosmetics, amounted to a whopping \$450 billion per year—and growing. The producers of these counterfeit goods, usually located in developing countries,

do not adopt sustainable, circular and ethical models, and cause great harm to European companies that are seriously committed to implementing them.

On the contrary, fashion and luxury consumers are becoming more and more demanding with regard to sustainability of the products they are buying; 66% of consumers are ready to pay more for products or services from companies committed to sustainability [16]. But sustainability is only possible when supported by production traceability, which demonstrates how and where the manufacturing process is carried out. In addition, in recent years, duties have been increased as the most evident aspect of international commercial turbulences. Since they are calculated on the basis of the Preferential Certification of Origin (PCO), a proper traceability of production is becoming mandatory to simplify the export procedures and to address the increasing requirements of custom agencies.

4.2 Supply Chain Track and Trace Process Value

Traceability by blockchain technology provides all the information to support informed purchase decisions of consumers, favouring real sustainable products. We apply blockchain technology to build a shared tamper-proof ledger that tracks the fabric manufacturing from source to sales. Our *sheep to shop track and trace blockchain*-based solution records the transformation of raw materials into fabrics and enables verification of EU PCO.

The expected impact is providing a complete and controlled set of information to support the efforts of the Piacenza company in the field of sustainability, environmental protection and ethical respect. The proposed solution leverages the competitive positioning of Piacenza and its customers, by providing final consumers with full provenance of items and documents. In addition, blockchain enables the full visibility of textile manufacturing by a safe and not modifiable process, which prevents the market from being affected by counterfeiting and unfair competition.

4.3 Distributed Ledger Implementation for Supply Chain Visibility

The blockchain solution implemented in the Piacenza trial records all steps and documents in the production process in a general way, storing documents hash on the ledger (on-chain) and a reference to their physical location while assuring their authenticity.

Figure 9 illustrates the main components of the supply chain visibility solution. Real data flows from Piacenza's ERP system through a wrapper so data can be written to the blockchain ledger through a RESTful API. The wrapper extracts the data from the ERP system in JSON format that matches the blockchain data model.

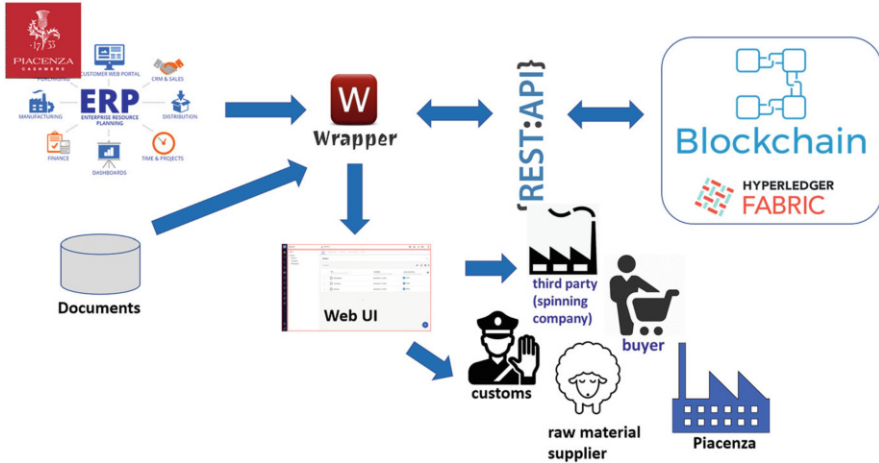


Fig. 9 High-level overview of the solution

The wrapper also feeds a dedicated Web UI whose role is to show the provenance of a specific selected item along with the corresponding documents as follows: The UI feeds the data from the API wrapper. The wrapper has a recursive function to retrieve every element in the chain. The recursive function calls the blockchain API to retrieve the information. The same information stored in the blockchain is then displayed in the UI. The PCO and other tracked documents information is displayed in the UI with a link to download the document. In other words, for a selected tracked item it graphically depicts its provenance and the (validated) documents stored on the ledger. This web UI can serve all participants in the network to trace a specific item and to check for specific documents (e.g., customs asking for a specific PCO).

Figure 10 shows the main modules for our sheep to shop blockchain application:

- **Blockchain infrastructure:** We apply Hyperledger Fabric [17] components. Our data model consists of two primary entities: *trackedItem* and *document*.
- **Smart contracts layer:** Smart contracts (chaincodes in Fabric) embed the business logic of the solution. Smart contracts functions are accessed through the Hyperledger Fabric Client (HFC) Software Development Kit (SDK) in Node.js.
 - Query functions enable accessing and fetching information stored in the ledger, including trackedItems and documents.
 - Invoke functions include the possibility of creating trackedItems and documents, and connecting a new document to an existing trackedItem.
 - Administration functions enable the management of the channels implemented as well as basic functions such as enrolment and registration.

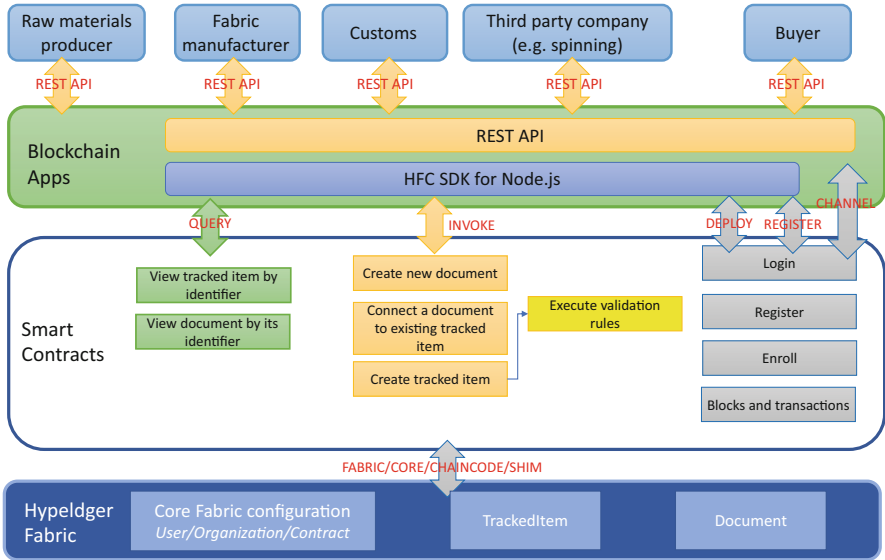


Fig. 10 Blockchain solution high-level design

- **Blockchain apps:** HFC SDK allows developing a blockchain client application which can use the SDK to invoke smart contract functions. This client can serve as a middle layer between frontend applications and the backend blockchain platform by providing RESTful APIs to be used by frontend applications.

4.4 Observations and Lessons Learned

Our achievements include the blockchain backend (released to open source under Apache v2 license [18]) and a (private) repository containing the dedicated code developed for extracting data from Piacenza’s ERP system and enabling the display of the provenance of items in the chain along with scripts, data and documents. The trial has emulated a complete blockchain business network. Obviously, the most natural way of extending and exploiting the results of the trial is by gradually incorporating Piacenza partners to the business network (e.g. customs and buyers) through the APIs provided. The provided blockchain backend is generic so this can be done in a straightforward manner.

The more challenging part is, therefore, not the technical but the business one, by defining a business model of onboarding, how to manage the network and how to monetize the savings and costs of participating and managing such a network. There are compelling evidences that show a great potential for this first trial solution to be extended to a full production environment for a full transparent and trackable solution towards a sustainable textile supply chain.

5 Conclusions

This chapter has introduced two large-scale trials that have been implemented in the context of the lighthouse project Boost 4.0. The Chapter has introduced the Boost 4.0 Reference Model, which adapts the more generic BDVA big data reference architectures to the needs of Industry 4.0. The Boost 4.0 reference model includes, on one hand, a reference architecture for design and implementation of advanced big data pipelines and, on the other hand, the digital factory service development reference architecture. Thus, Boost 4.0 can fully address ISO 20547 for Industry 4.0.

This chapter has demonstrated that the BDVA big data reference architecture can indeed be adapted to the needs of the Industry 4.0 and aligned with an overall digital factory reference architecture, where big data-driven processes will have to extend advanced manufacturing processes such as smart engineering, smart planning and commissioning, smart workplaces and operations, smart connected production and smart maintenance and customer services. Such digital factory service development architecture can indeed host and accommodate the needs of advanced big data-driven engineering services.

The chapter has demonstrated that both intra-logistic process planning and connected supply chain track and tracing can achieve significant gains and extract significant value from the deployment of big data-driven technologies. The evolution from traditional data analytic architecture into big data architectures will enable increased automation in simulation and process optimization. The combination of Industry 4.0 data models such as OPC-UA, AML and IoT open APIs such as FIWARE NGSI allows for dynamic and real-time optimization of intra-logistic processes compared to off-the-shelf commercial solutions. Moreover, big data architectures allow a higher granularity and larger simulation scenario assessment for a high-fidelity intra-logistic process commissioning.

The use of open-source big data technology suffices to meet the challenge of very demanding big data processes in terms of variety, velocity and volume as the VWAE trial has demonstrated. This trial has also shown that digital twin operations can be greatly improved if supported by advanced big data streaming technologies, and the use of shared data spaces demonstrates the suitability of such technologies to break information silos and increase efficiency and scale up intralogistics processes. This chapter has also shown that distributed ledger technology can be seamlessly integrated with distributed data spaces and support business network traceability and visibility in the high-end textile sector (Piacenza trial). The chapter has also provided evidence on how the extensive use of open technologies, APIs and international standards can greatly support the large-scale adoption and uptake of big data technologies across large ecosystems. The chapter has provided compelling evidences that big data can greatly improve performance of Industry 4.0 engineering services, particularly when development and exploitation of digital threads and digital twins come into operation. The interested reader is also referred and invited to browse the content in chapters “Next Generation Big Data

Driven Factory 4.0 Operations and Optimisation: The Boost 4.0 Experience” and “Model Based Engineering and Semantic Interoperability for Trusted Digital Twins Big Data Connection Across the Product Life Cycle” focused on the Boost 4.0 lighthouse project; these chapters discuss how further trials have incorporated big data technologies as part of the business processes for increased competitiveness.

This research is opening the ground for implementation of more intelligent, i.e. cognitive and autonomous, intra-logistic processes. As the diversity of parts considered and the autonomy in decision process increase, further research is needed in terms of development of sovereign and large-scale distributed data spaces that can provide access to the necessary data for AI model training beyond pure data analytics and digital twin simulation. The Boost 4.0 big data framework calls for further research on the development of federated learning models that can combine highly tailored models matching and optimized to the specificities of the factory layout with more general models that can be shared and work at higher levels of abstractions; thus, speed and long-term planning can be combined in new forms of autonomous shopfloor and supply chain operations.

Acknowledgements This research work has been performed in the framework of the BOOST 4.0 Big Data Lighthouse initiative, a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 780732. This data-driven digital transformation research is also endorsed by the Digital Factory Alliance (DFA) www.digitalfactoryalliance.eu

References

1. Boost 4.0 <https://www.boost40.eu>
2. Zillner, S., Curry, E., Metzger, A., Auer, S., & Seidl, R. (Eds.). (2017). *European Big Data Value Strategic Research & Innovation Agenda*. Big Data Value Association.
3. Zillner, S., Bisset, D., Milano, M., Curry, E., García Robles, A., Hahn, T., Irgens, M., Lafrenz, R., Liepert, B., O’Sullivan, B., & Smeulders, A. (Eds.) (2020) *Strategic Research, Innovation and Deployment Agenda - AI, Data and Robotics Partnership. Third Release*. September 2020, Brussels. BDVA, euRobotics, ELLIS, EurAI and CLAIRE.
4. Big Data Value Association. (2017). *Strategic Research and Innovation Agenda v.4.0*. Available at: https://www.bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf
5. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik. (2015). *Status Report: Reference Architecture Model Industrie 4.0 (RAMI4.0) (Vol. 0)*. <https://www.zvei.org/Downloads/Automation/5305>. Publikation GMA Status Report ZVEI Reference Architecture Model.pdf.
6. Digital Factory Alliance (DFA). <https://digitalfactoryalliance.eu/>
7. *D2.5 boost 4.0 reference architecture specification*. Available at: <https://boost40.eu/wp-content/uploads/2020/11/D2.5.pdf>
8. International Organization for Standardization. (2011). *ISO/IEC/IEEE 42010. System and Software Engineering – Architecture Description*. Available at: <https://www.iso.org/standard/50508.html>
9. *Apache Spark*. <https://spark.apache.org/>
10. International Data Spaces Association. (2019). *Reference architecture model*. Online. Available at: <https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf>
11. *PostgreSQL*. <https://www.postgresql.org/>

12. *MongoDB*. <https://www.mongodb.com/>
13. *Apache Hive*. <https://hive.apache.org/>
14. *Grafana*. <https://grafana.com/>
15. Center for Trade Facilitation and Electronic Business, Economic Commission for Europe, UN. (2019). *Briefing note on sustainable textile value chains in the garment and footwear domain For SDG12*. [online] Geneva. Available at: <https://undocs.org/pdf?symbol=en/ECE/TRADE/C/CEFACT/2019/26>
16. Center for Trade Facilitation and Electronic Business, Economic Commission for Europe, UN. (2019). *Report of the Centre for Trade Facilitation and Electronic Business on its twenty-fifth session*. [online] Geneva. Available at: https://www.unece.org/fileadmin/DAM/cefact/cf_plenary/2019_plenary/ECE_TRADE_C_CEFACT_2019_02E_Report.pdf
17. NIELSEN. (2015). *The sustainability imperative. New insights on consumer expectations*. Online. Available at: <https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/global-sustainability-report-oct-2015.pdf>
18. *Hyperledger Fabric*. <https://www.hyperledger.org/use/fabric>
19. Available at: <https://gitlab.com/boost4-piacenza/public-artifacts>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

