BIG DATA FOR FACTORIES

Big Data Value Spaces for Competitiveness of European Connected Smart Factories 4.0

Horizon 2020 EU Grant Agreement 780732

Title	D2.5 – BOOST 4.0 Reference Architecture Specification v1
Document Owners	INTRA
Contributors	VIS, ASTI, TID, UNINO, TTT, RISC, PEN, IMEC, SIEMENS, PRIMA, POLIMI, I2CAT, TRIMEK, CAPVIDIA, CHAL, SAS, FhG, ATLANTIS, EPFL, ATB, UBO, CERTH, UED, IMT, IDSA, FF, IBM, ESI, ENEO, SQS, CARSA, UTC, UNP, LUH
Dissemination	Public
Date	07/01/2019
Version	V1.0

Version history

09/11/2018	Table of contents.
15/11/2018	First INTRA contribution.
20/11/2018	Fist content review from internal.
30/11/2018	Partner input compiled, ready for review.
14/12/2018	Updated document based on review comments, ready for second review round.
20/12/2018	Updated to resolve second review comments and to include missing parts.
07/01/2019	Added ERCIM contribution and addressed comments.

Document History

Authors	VIS, ASTI, TID, UNINO, TTT, RISC, PEN, IMEC, SIEMENS, PRIMA, POLIMI, I2CAT, TRIMEK, CAPVIDIA, CHAL, SAS, FhG, ATLANTIS, EPFL, ATB, UBO, CERTH, UED, IMT, IDSA, FF, IBM, ESI, ENEO, SQS, CARSA, UTC, UNP, LUH
Internal Reviewers	EPCC, TID
Workpackage	WP2
Task	T2.4
Nature	Report
Dissemination	Public

Project Partners

Participant organisation name	Acronym
Asociación de Empresas Tecnológicas Innovalia	INNO
Volkswagen Autoeuropa, Lda *	VWAE
Visual Components	VIS
Automatismos y Sistemas de Transporte Interno S.A.U.	ASTI
Telefónica Investigación y Desarrollo SA	TID
Volkswagen AG. *	VW
UNINOVA	UNINO
FILL GmbH. *	FILL
TTTECH Computertechnik AG	TTT
RISC Software GmbH	RISC
PHILIPS Consumer Lifestyle B.V. *	PCL
PHILIPS Electronics Nederland	PEN
Interuniversitair Micro-Electronicacentrum VZW	IMEC
Centro Ricerche Fiat S.C.p.A. *	CRF
SIEMENS S.p.A.	SIEMENS
Prima Industries S.p.A	PRIMA
Politecnico di Milano	POLIMI
AUTOTECH ENGINEERING, AIE *	GESTAMP
Fundació Privada I2CAT, Internet I Innovació Digital A Catalunyai2cat	I2CAT
TRIMEK S.A.	TRIMEK
CAPVIDIA N.V,	CAPVIDIA
Volvo Lastvagnar AB *	VOLVO
Chalmers Tekniska Hoegskola AB	CHAL
Whirlpool EMEA SpA *	WHIR
SAS Institute Srl	SAS
Benteler Automotive GmbH *	BAT
It.s OWL Clustermanagement	OWL
Fraunhofer Gesellschaft Zur Foerderung Der Angewandten Forschung E.V.	FhG
Atlantis Engineering	AE
Agie Charmilles New Technologies SA *	+GF+
Ecole Polytechnique Federale De Lausanne	EPFL
Institut Für Angewandte Systemtechnik Bremen GmbH	ATB
Rheinische Friedrich-Wilhelms-Universitat Bonn	UBO
Ethniko Kentro Erevnas Kai Technologikis Anaptyxis (CERTH)	CERTH

The University of Edinburgh	UED
Institute Mines Telecom	IMT
International Data Spaces e.V.	IDSA
FIWARE Foundation e.V	FF
GEIE ERCIM EEIG	ERCIM
IBM ISRAEL – Science and Technology LTD	IBM
ESI Group	ESI
Eneo Tecnología, S.L	ENEO
Software Quality Systems S.A.	SQS
Consultores de Automatización y Robótica S.A.	CARSA
INTRASOFT International	INTRA
United Technologies Research Centre Ireland, Ltd *	UTRC-I
Fratelli Piacenza S.p.A. *	PIA
RiaStone - Vista Alegre Atlantis SA *	RIA
Unparallel Innovation, Lda	UNP
Gottfried Wilhelm Leibniz Universität Hannover	LUH

*LHF 4.0 – Lighthouse Factory 4.0   * RF – Replication Factory 4.0

Executive Summary

This document details the Boost 4.0 Reference Architecture (RA) v1, which serves a common point of understanding between Boost 4.0 partners; encourages adherence to common standards, specifications, and patterns; illustrates and improves understanding of the various Big Data components, processes and systems; and supports communication with other communities developing Reference Architectures for the manufacturing domain.

In this context, the elements of the Boost 4.0 RA are detailed in paragraphs 2 to 4. The following figure provides an overview of the architecture:

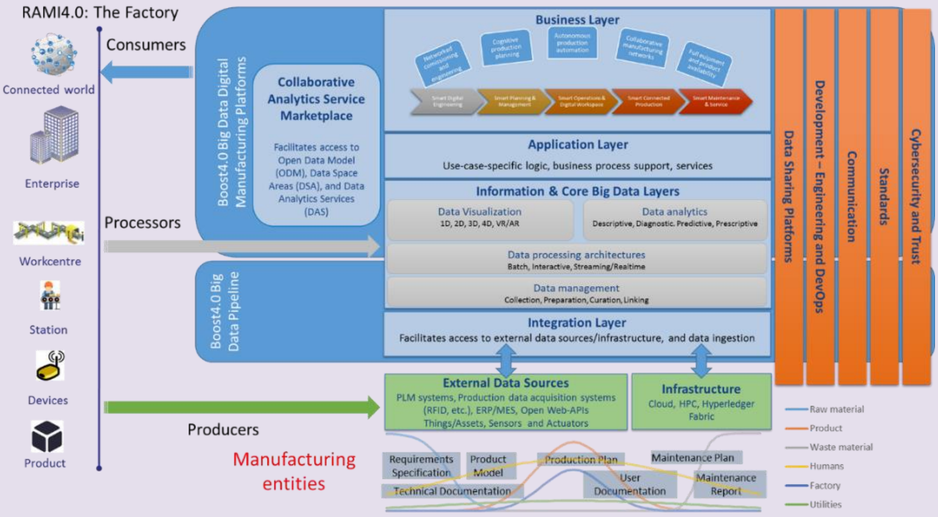

Figure: Overview of the Boost4.0 RA

It consists of the following horizontal layers grouping together various components:

- **Integration layer**, which facilitates management of external data sources and of infrastructure, as well as data ingestion.
- **Information and Core Big Data layers**, consisting of components responsible for data management, processing, analytics and visualisation.
- **Application layer**, which represents components implementing application logic that supports specific business functionalities and exposes the functionality of lower layers through appropriate services.
- **Business layer**, which form the overall manufacturing business solution in the Boost 4.0 project.

The horizontal layers are supplemented by the cross-cutting aspects, the factory dimension and the manufacturing entities dimension:

- **Communication**, which provides the mechanisms and technologies for the reliable transmission, and receipt of data.
- **Data sharing platforms**, which considers the needs for allowing data providers to share their data as a commodity.
- **Development- Engineering and DevOps**, which covers tool chains and frameworks that significantly increase productivity in terms of developing and deploying big data solutions.
- **Standards**, which elaborates on the standards use within the Boost4.0 RA and connect to the work documented in D2.7.
- **Cybersecurity and Trust**, which covers topics such as device and application registration, identity and access management, data governance, data protection, and so on.
- **The Factory dimension** indicates how physical entities are engaged in the RA.

- **The manufacturing entities dimension** allows the specification of elements in the manufacturing domain across the manufacturing process life-cycle stages and adds the relevant data types and ontologies for them.

Section 5 deals with the alignment with existing reference architectures in the domains of Big Data and manufacturing. It details how BDVA Big Data Value Reference Model, NIST Big Data Reference Architecture, RAMI4.0, and IIRA layers, verticals, and other aspects correspond to the parts of the Boost4.0 detailed in sections 2-4. Additionally, this section maps the Boost4.0 RA to the Digital Shopfloor Alliance Data-driven Smart Solution Development Framework, and finally how Boost 4.0 RA supports the EFFRA Autonomous Factory pathway in terms of Digital Architecture.

Section 6 details the mapping of major Boost 4.0 frameworks: these are Industrial/International Data Spaces (IDS), Big Data Europe (BDE), and Hyperledger Fabric. For IDS and BDE, a clear mapping between their internal architectures and the Boost 4.0 RA.

Section 7 and 8 details the solutions of Boost4.0 within the various layers of the RA and detail how these solutions are applied in the project pilots.

In conclusion, the BOOST4.0 refence architecture:

- Bridges together the Big Data domain with the manufacturing by aligning well with existing models and architectures in both of them.
- Supports the pilot in identifying big data assets and promotes the understanding of how major frameworks/technologies in the project may be applied.
- Supports the capability of mixing together open source and proprietary components for enabling pilot solutions.

Keywords: Big Data, Reference Architecture, Layered Architecture, Factory dimension, Data Dimension, Manufacturing.

Disclaimer

Acknowledgement

This document is a deliverable of Boost 4.0 project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 780732.

Table of contents

Abbreviations and Acronyms

Acronym	Meaning
A.I.	Artificial Intelligence
API	Application Programming interface
APM	Application performance management
BDE	Big Data Europe
BDVA	Bid Data Value Association
CI	Continuous Integration
CPS	Cyber Physical System
CRM	Customer Relationship Management
DevOps	A clipped compound of "development" and "operations
ERP	Enterprise Resource Planning
ETL/ELT	Extract, Transform, Load / Extract, Load, Transform
HMI	Human Machine Interface
HPC	High Performance Computing
HRM	Human Resource Management
IDS	Industrial/International Data Spaces
IIC	Industrial Intern Consortium
IIRA	Industrial Internet Reference Architecture
IIoT	Industrial Internet of Things
IoT	Internet of Things
IT	Information Technology
MES	Manufacturing Execution System
NBDRA	NIST Big Data Reference Architecture
NRT	Near Real Time
PLM	Production Lifecycle Management
RA	Reference Architecture
RAMI4.0	Reference Architecture Model Indusrtie 4.0
RFID	Radio-Frequency Identification
IDS	International Data Spaces

List of Figures

List of Tables

1 Introduction

According to Reference Architectures and Open Group Standards for the Internet of Things [1] a reference architecture is a generic architecture that provides guidelines and options for making decisions in the development of more specific architectures and the implementation of solutions.

The goal of the Boost 4.0 Reference Architecture (RA) is to develop a reference architecture for Big Data in the manufacturing domain that achieves the following objectives:

1. Provides a common language for the various stakeholders.
2. Encourages adherence to common standards, specifications, and patterns.
3. Illustrates and improves understanding of the various Big Data components, processes, and systems.
4. Supports communication with other communities developing Reference Architectures for the manufacturing domain (for example RAMI4.0, IIRA).

The Boost 4.0 RA also serves as a tool to facilitate discussion of the requirements, design structures, and operations inherent in Big Data and is intended to facilitate the understanding of the operational intricacies in Big Data in context of the manufacturing domain. It does not represent the system architecture of a specific Big Data system, but rather is a tool for describing, discussing, and developing system-specific architectures using a common framework of reference. The model is not tied to any specific vendor products, services, and so forth. Furthermore, it is expected that with the complexity of the Boost4.0 project, which encompasses 10 pilot cases, the RA presented in this document will allow for flexibility in implementing solutions that partially address the RA's aspects.

2 Boost 4.0 Reference Architecture Overview

The Boost 4.0 RA consists of a number of layers at its core, alongside with a Factory dimension and a manufacturing entities dimension.

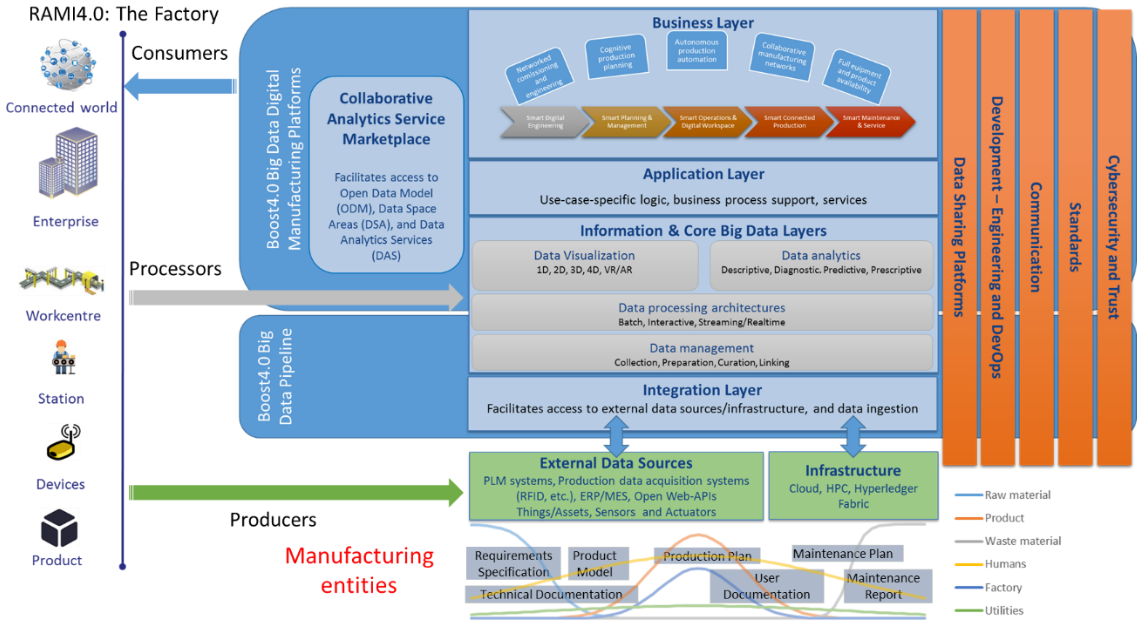

Figure 1 The Boost 4.0 RA

The core horizontal layers represent a collection of functionalities/components performing a specific role in the data processing chain, and are:

1. Integration layer, which facilitates management of external data sources and of infrastructure, as well as data ingestion.
2. Information and Core Big Data layers, consisting of components responsible for data management, processing, analytics and visualisation.
3. Application layer, which represents components implementing application logic that supports specific business functionalities and exposes the functionality of lower layers through appropriate services.
4. Business layer, which form the overall manufacturing business solution in the Boost 4.0 five domains (networked commissioning & engineering, cognitive production planning, autonomous production automation, collaborative manufacturing networks, and full equipment & product availability) across five process life-cycle stages (Smart Digital Engineering, Smart Production Planning & Management, Smart Operations & Digital Workplace, Smart Connected Production, Smart Maintenance & Service).

Outside the aforementioned core horizontal layers, the Collaborative Analytics Service Marketplace has been added and facilitates the use of data and services by interested parties. This element will interface with the components belonging to the various layers in order to achieve its purpose.

Apart from the horizontal layers, there are a number of vertical cross-cutting aspects that affect all layers:

1. **Communication**

   This aspect aims to provide the mechanisms and technologies for the reliable transmission, and receipt of data between the layers of the Boost4.0 RA.
2. **Data sharing platforms**

   This aspect considers the needs for allowing data providers to share their data as a commodity, covering specific data, for a predefined space of time, and with a guarantee of reversibility at the end of the contract.
3. **Development- Engineering and DevOps**

   This aspect covers tool chains and frameworks that significantly increase productivity in terms of developing and deploying big data solutions.
4. **Standards**

   This section will elaborate on the standards use within the Boost4.0 RA and connect to the work documented in D2.7.
5. **Cybersecurity and Trust**

   This aspect covers topics such as device and application registration, identity and access management, data governance, data protection, and so on

Additionally, the Boost 4.0 RA considers the external data sources that may include PLM systems, Production data acquisition systems, MES/ERP systems, Open web APIs, and sensors/actuators at the shop floor.

The RA considers the Factory (as defined in RAMI4.0) dimension to indicate how physical entities are engaged in the RA. Finally, the dimension of manufacturing entities allows the specification of such elements across the manufacturing process life-cycle stages and adds the relevant data types and ontologies for them.

3 Reference Architecture Layers and Dimensions

3.1 Integration layer

The integration layer groups together components for facilitating access to external data sources such as PLM systems, production data acquisition systems, MES/ERP systems, Open web APIs, and sensors/actuators. In the case where controlling (or providing feedback to) shop floor assets is required, the functionality to realise this type of communication is also provided by components belonging to this layer. Integration layer also includes components that facilitate access to Boost4.0 infrastructure such as cloud, HPC, and Hyperledger Fabric. The components in this example typically follow a service provider-consumer paradigm where a 'facilitator' component is used to enable interoperability in terms of the data exchanged or the interfaces provided. Other options include the use of a common data model, or internal data transformation within the components to make the exchanged data compatible with their internal data models.

Connectivity between the Integration layer and compute and data components is dependent upon several factors:

1) Type of external data sources in use -for example a traditional file system; object store; relational databases management systems or other databases such as NoSQL databases; high velocity data such as sensor data.
2) Type of interfaces offered by any compute resources. An HPC system may not expose its service with an API; and as such bespoke connectors would need to be created. Some cloud providers, such as Microsoft Azure and Amazon AWS, do offer a programmable interface to allow job submission via some web-portal.
3) The types of APIs offered by any of the listed. Programmable interfaces are likely not to be compatible, and thus software engineering effort would be required to ensure compatibility. Whereas Object Stores have de facto standard based API such as S3 interface others don't have such standards to access them.

3.2Information and Core Big Data layers

The Information and Core Big Data Layers consist of components that belong in the following sub-layers:

1. **Data management**

   This layer groups together components facilitating data collection, preparation, curation, and linking.

2. **Data processing architectures**

   This layer groups together components to support implementation of applications that can deal with the volume, velocity, variety and variability of data, focusing on architectures for data manipulation.

3. **Data analytics**

   This layer groups together components to support data analysis covering descriptive, diagnostic, predictive, and prescriptive data analysis.

4. **Data visualisation**

   This layer groups together algorithms/components to support data visualisation and user interaction.

The following figure highlights these layers:

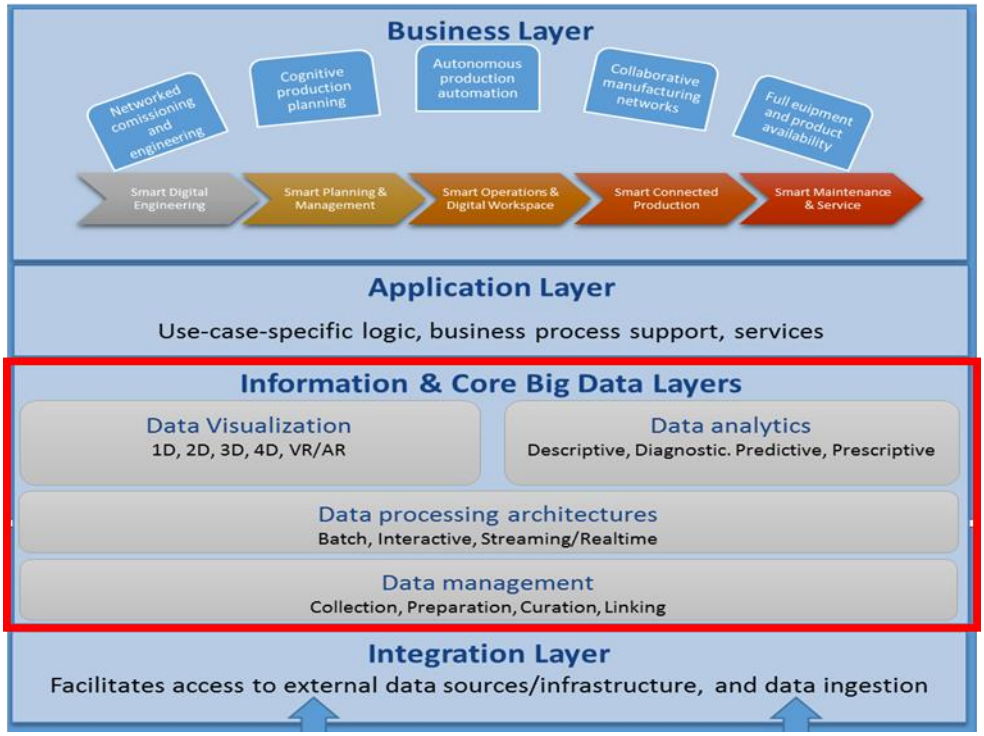

Figure 2 Information & Core Big data Layers in the RA

It should be noted that this layering does not imply that each layer needs to build on top of the previous one – that is, a component in some layer may directly interact with a component in some layer further below, for example, a visualisation tool may get data directly from a data source.

Additional details for each of these sub-layers are provided in the sections below.

3.2.1 Data management

3.2.1.1 Collection

In general, the collection activity handles the interface with the Integration Layer components. This may be a general service, such as a file transfer server (FTP) or web server configured to accept or perform specific collections of data, or it may be an application-specific service designed to pull data or receive data pushes. Since this activity is receiving data at a minimum, it must store/buffer the received data until it is persisted. Persistence may be to an in-memory queue or equivalent service, or to physical media. The collection activity is likely where the extraction portion of the Extract, Transform, Load (ETL)/Extract, Load, Transform (ELT) [89] cycle is performed. At the initial collection stage, sets of data (for example data records) of similar structure are collected in their original form. Initial metadata is created (for example subjects with keys are identified) to facilitate subsequent aggregation or look-up methods.

3.2.1.2 Preparation and Curation

The preparation and curation activities convert the collected data into cleansed organised data and in forms that can be effectively processed and analysed. This includes producing customised data that can be processed in a business solution's pipeline of big data components. Tasks performed include data validation (for example, checksums/hashes, and format checks), cleansing (for example, eliminating bad records/fields), outlier removal, standardisation, reformatting and encapsulation. Such tasks are frequently executed where source data will be persisted and provenance data will be verified or attached/associated. Verification or attachment may include optimisation of data through manipulations (for example deduplication) and indexing to optimise the analytics process. This activity may also aggregate data from different sources, leveraging metadata keys to create an expanded and enhanced data set.

Data curation requires active management throughout the data lifecycle. According to the DCC model [49] this lifecycle comprises of conceptualising, creating, accessing, using, appraising, selecting, disposing, ingesting, reappraising, storing, reusing, and transforming data. During this process, data might be annotated, tagged, presented, and published for various purposes. Data curation helps reduce threats to data long-term value and mitigate digital obsolescence.

Additionally, these tasks need to be complemented with data governance. According to DAMA International Data Management Book of Knowledge, "Data Governance is defined as the exercise of authority and control (planning, monitoring and enforcement) of data assets." Implementation of data governance results in policies on how to handle data. Data curation may make use of data governance when customising information.

Data curation provides the following benefits according to [2]:

- Increases Machine Learning effectiveness
- Avoid the Data Swamps issue: A Data Lake strategy allows users to easily access raw data, to consider multiple data attributes at once, and the flexibility to ask ambiguous business driven questions. But Data Lakes can end up Data Swamps where finding business value becomes complicated (for example impossible to analyse data, data is used inappropriately because of inaccuracies).
- Ensures Data Quality: Data curation foresees data cleaning and actions to ensure the long -term preservation and retention of the authoritative nature of digital objects.

3.2.1.3 Linking

The data linking process combines the data from different sets into one population or multiple linked ones, matching, aligning and integrating information as well as making sure that duplicates are eliminated. Usually the following methods are used to facilitate linking:

1. Deterministic

Deterministic linkage is based on matching records from different data sets based on one or more reliable identifiers. The high degree of certainty required for deterministic linkage is achieved through the existence of a unique identifier for a record, such as company ID number. Combinations of variables (for example first name, last name, sex) may also be used as unique identifiers.

2. Probabilistic

Probabilistic linkage is used when there are no unique record identifiers, or when the variables used are not as accurate, stable or complete in order to achieve deterministic linkage. In this case linkage depends on achieving a close approximation by a using a wider range of identifiers. Probabilistic linkage has a greater capacity to link when errors exist in linking variables, so it may lead to much better linkage than simple deterministic methods.

Both deterministic and probabilistic procedures should be considered iterative. After completing the initial linkage, a random sample of match decisions should be reviewed to ensure that the algorithm is performing as intended.

3.2.2 Data processing architectures

The Data processing architectures for Big Data provide the software infrastructure to support implementation of applications that deal with the volume, velocity, variety and variability of big data. Data processing architectures define how data processing is handled, and therefore are categorised based on whether they support on "batch" and "stream" processing. However, recently also hybrid processing systems have emerged. The following paragraphs will provide an overview of the most commonly used solutions.

3.2.2.1 Batch processing

When the batch processing model is used, a large volume of data (usually collected over a period) is fed into an analytics system. An example of such a data set are factory legacy systems that have accumulated data for extended periods of time. Batch processing works well in situations where real-time analytics results aren't required, and large volumes of data have to be processes in order to produce the analytics results.

Apache Hadoop

Apache Hadoop [90] is an open-source software framework for distributed storage and batch processing of extremely large data sets. It relies on MapReduce, a programming model and implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster. According to [50] Apache Hadoop provides a scalable, cost effective, resilient, and highly available solution for batch processing. However, the following key limitations have been identified in [51]:

1. It is not suited for small data.
2. By default, it does not support real-time data processing.
3. It ensures that data job is complete, but it is unable to guarantee when the job will be complete.
4. Security limitations.

3.2.2.2 Stream processing

Contrary to the batch processing model, under the stream processing model data is fed into analytics systems piece by piece. This approach works well with sensor data deployed in the factory shop floor and can produce real time or near real time results.

Apache Storm

Apache Storm [90] is an open source distributed real-time computation system. Storm makes it easy to reliably process unbounded streams of data, doing for real-time processing. Its advantages include scalability, low latency, reliability (guarantees that each unit of data will be processed at least once or exactly once), and fault-tolerance.

Apache Samza

Apache Samza [92] is an open-source near-real-time, asynchronous computational framework for stream processing. Its advantages include scalability, low latency, and fault-tolerance.

Key differences

According to [52] key differences include:

1. Both offer at least once delivery for messages (duplicates may be received). But Apache Storm is capable of exactly-once using Trident.
2. Apache Samza, supports Java Virtual Machine-based languages only. While Apache Strom also supports Python.
3. Apache Samza delivers higher throughput.

3.2.2.3 Hybrid Processing Systems: Batch and Stream Processors

Apache Spark and Apache Flink are the major solutions capable of processing both batch and stream workloads.

Apache Spark

Apache Spark [93] is an in-memory batch data processing framework and supports stream processing by micro-batching. Spark supports not only MapReduce operations but also iterative operations like in machine learning, stream processing, graph data processing and interactive queries. Spark offers increased performance when compared to Apache Hadoop due to its in-memory processing. In addition, it provides scalability, and fault tolerance.

Apache Flink

Apache Flink [94] is one of the more recent stream processing frameworks that can process streaming data as well as batch data. At its core Apache Flink is stream processing framework capable of handling batch processing as special case of streaming with bounded data. Its advantages include scalability, low latency, and fault-tolerance. As a relatively new framework Flink has less APIs and a smaller base of contributors when compared to Apache Spark.

3.2.3 Data analytics

Data analytics components can be divided in the following categories based on their functionality:

1. Descriptive
2. Diagnostic
3. Predictive
4. Prescriptive

Descriptive analysis uses past (events that have occurred) data to provide results interpretable by humans. The majority of statistics belong to this category. Examples include reports that provide historical insights regarding a company's production, operations, and sales and so on.

Diagnostic analytics use data to uncover the root cause behind an incident and may employ machine learning techniques to recognise patterns, detect anomalies or 'unusual' events, and identify drivers of KPIs.

Predictive analytics are about understanding what might happen in the future, and provide insights based on past data. They combine historical data to identify patterns and apply statistical models and algorithms to capture relationships between various data sets.

Prescriptive analytics attempt to predict the effect of future decisions in order to support decision making on possible outcomes before any decisions are made. Prescriptive analytics use a combination of techniques and tools like machine learning and computational modelling procedures on various input from different data sets (historical data, real-time data, and so on).

3.2.4 Data visualisation

Big Data visualisation involves the presentation of data of almost any type in a graphical format that makes it easy to understand and interpret, enabling decision makers to explore data sets to identify correlations or unexpected patterns.

Big Data visualisation relies on powerful computer systems to ingest raw corporate data and process it to generate graphical representations that will allow users to take in and understand and interpret vast amounts of data.

Big Data visualisation techniques include:

1. **Linear**: Lists of items, items sorted by a single feature.
2. **2D/Planar/geospatial:** Cartograms, dot distribution maps, proportional symbol maps and contour maps.
3. **3D/Volumetric:** 3D computer models, computer simulations
4. **Temporal:** Timelines, time series charts, connected scatter plots, arc diagrams, circumplex charts.
5. **Multidimensional:** Pie charts, histograms, tag clouds, bar charts, tree maps, heat maps, spider charts.
6. **Tree/hierarchical:** Dendrograms, radial tree charts and hyperbolic tree charts.

3.3 Application layer

The application layer represents a group of components that supports specific business functionalities and exposes the functionality of lower layers through appropriate services. Components in this layer apply application logic, rules and models at a coarse-grained, high level, and rely on the components of the layer below for lower level operations.

The number of services deployed in this layer is not restricted therefore allowing the decomposition of larger applications into several small services, thus improving modularity of the final solution, and realising a service-oriented approach relying on micro-services.

Additionally, this layer supports the realisation of the Digital Shopfloor Alliance Data-driven Smart Solution Development Framework which foresees a layered approach at different levels (Enterprise, Factory, Workcell / Production Line and Field Devices). On each layer/level, different tools or services are applied or offered, and for all of them different modelling approaches may be used. According to the Boost 4.0 RA, the DSA services for each level could be deployed within the application layer supporting different business functionalities in the Business layer but also enabling the communication with each other, thus enabling the implementation of new services based on the lower level DSA services.

3.4 Business layer

The business layer combines the functionalities of the layers below to realise and link different business processes in support of the business models of each pilot. The identified process life-cycle stages are:

- Smart Digital Engineering
  Collaborative engineering for redesign of the manufacturing system and product considering feedback from data acquired.
- Smart Planning & Management
  Digital factory modelling and simulation, including access through cloud to modelling for process improvement and control, for example using machine learning methods.
- Smart Operations & Digital Workspace
  Multiple source data mining and real time advanced analytics at the Factory and Value Network Levels.
- Smart Connected Production
  Manufacturing as a Service (MaaS) for digital twin manufacturing capability extension across the entire value chain.

- **Smart Maintenance & Service**
  Servitisation of maintenance, where data is shared between manufacturer and maintenance supplier.

The supported business solution domains are:

- **Networked commissioning and engineering**
  Short time-to-market of innovative customised products is a key success factor for industrial companies. Integrating big data feedback information from operation and maintenance phases into the engineering phases will shorten the time for real plant or factory commissioning.
- **Cognitive production planning**
  Allows system wide visibility and data flow and analysis from the shop-floor to the top floor to the global value chain will allow manufacturing companies to trade on their production capacity, manufacturing planning schedules and production costs to perform dynamic end-to-end production planning across flexible value networks.
- **Autonomous production automation**
  Multi-source high-speed production data processing in workplace-process-human-machine context evaluation is critical for shop-floor productivity and safety concerns. Machine learning based high-performance production data analysis is key for system autonomy-automation and augmented human competences.
- **Collaborative manufacturing networks**
  Hundreds of thousands of parts, provided by multiple supplier facilities across the globe, go into large complex product such as automotive or smart home appliances. Big data connects physical production world with the digital twin. Big data transparency means continuous process coordination and enables quality control within and across the complete value chain.
- **Full equipment and product availability**
  Continuous product or machine data means continuous analysis, risk assessment, and process coordination resulting in better customer experience, fewer field service calls, optimum spare part distribution and prescriptive maintenance.

3.5The factory dimension

The RAMI4.0 factory dimension [3] is used to identify the roles of the various manufacturing entities in context of the RA. These roles are Producers, Consumers and Processors.

1 Producers feed information into the Boost 4.0 RA.

2 Consumers use the service, tools, and/or data offered by the Boost 4.0 platform.
3 Processors are the elements hosting the components of the Boost.4.0 platform.

The producer elements help identify the physical location of the data-sources and supporting the deployment of the components in the integration layer.

The consumer elements identify the elements targeted by the components in the business/application layers. These elements require access specific components and therefore affect the deployment and even the implementation of these components.

The Processors elements identifies those elements that host specific components of the Boost 4.0 RA, thus facilitating their deployment.

3.6Manufacturing entities dimension

In the reference architecture, the main effort towards standardisation will be laid on the design of the manufacturing entities dimension, and the establishment of the reference architecture as a means to provide standards-based access to all the data in the manufacturing domain of interest. The manufacturing entities dimension deals with the design and implementation of the semantic model and the mechanisms for intelligent filtering. It will provide a unified and all-spanning semantic model covering the multi-domain knowledge of the BOOST 4.0 pilots. Therefore, the reference architecture will achieve semantic enrichment through manufacturing entities dimension which deals with cross-sectoral knowledge exploration and filtering. Thus, this dimension requires semantic interoperability, that is, the ability of information systems to exchange data unambiguously with a shared meaning as a standard of all the BOOST 4.0 pilots.

Standardisation activities will be driven through involvement with the Industry Ontology Foundry (IOF) [118] group and regular participation in their workshops. IOF is an international foundry aiming at:

i) Providing principles and best practices to develop quality ontologies that will support interoperability for industrial domains.
ii) Creating a suite of open and principles-based ontologies from which other sub-domain or application ontologies can be derived in a modular fashion, remaining 'generic' (non-proprietary, non-implementation specific).
iii) Instituting a governance mechanism to maintain and promulgate these goals and principles.
iv) Providing an organisational framework and governance processes that ensure conformance with the principles and best practices for development, sharing, maintenance, evolution, and documentation of IOF ontologies.

The purpose of these reference ontologies is to allow for extensions to be progressive to more specific or constrained sub-domains. To meet this goal, the IOF ontologies are expected to have an architecture that starts from alignment with a domain neutral ontology, also referred to as an Upper Ontology or Foundational Ontology, from which subsequent IOF ontologies can be developed (newly or adapted from existing ones) that are ontologically consistent, coherent and modular, allowing for reusability. BOOST 4.0 pilots will provide testbeds to build a reference semantic model in the manufacturing domain (contributing pilot-specific terminology and domain knowledge in a bottom-up way, and then linking it to suitable existing, standard-based semantic models), and IOF will provide methodologies and principles for implementation of a standard semantic model (a semantic model of the overall BOOST 4.0 architecture, addressing cross-pilot requirements). Some of the required semantic models will not emerge from scratch in this process but have been initiated before BOOST 4.0. Whenever such models require further maturing, the partners agree to feed these models into the IOF process to improve their quality together with the IOF group. To summarise, the main objectives of the manufacturing entities dimension are:

- To identify the domain of interest, covering all relevant products, data sources, information flow resources, design and manufacturing processes, user-interface access points and dynamics of the entire system.
- To design and implement the semantic model.
- To provide a linked data integration framework that will extract, export, and harmonise data from various sources.
- To enable semantic enrichment (for example, annotations, tagging) of data originating from disparate research or existing systems (for example, PLM/CAx).
- To provide an intelligent engine that will allow active exploration of the linked data sets and implicit knowledge discovery.

4 Cross-cutting aspects

4.1 Data sharing platforms

This section contains an overview of the following relevant data sharing platforms: Unified eXchange Platform, International/Industrial Data Spaces, Linked (meta) data pipeline/Open Micka and the European Open Science Cloud.

4.1.1 Unified eXchange Platform

Unified eXchange Platform (UXP) [53] is a technology that enables peer-to-peer data exchange over encrypted and mutually authenticated channels. It is based on a decentralised architecture where each peer has an information system that will be connected with other peers' systems.

The Unified eXchange Platform (UXP) is targeted at situations where several parties wish to establish a standardised communication channel that provides confidentiality, strong authentication and long-term proof value of the relayed messages. In the simplest case, the system has the following participants:

1. Members – entities that wish to communicate with each other. The assumption is that each member has an information system that will be connected with other members' systems
2. Governing agency – coordinates communication activities, creates and distributes security policy, maintains and distributes registry of members, distributes gateway software.
3. Trust service providers – provide certification and time-stamping services. In the simple case, the trust services can be provided by the governing agency.

In the UXP the communication is organised as synchronous service calls. The service providers design and implement services and make them available for service clients. Access to the services is controlled by service provider. In order to use a service, the service client and the service provider enter into agreement that specifies the terms of the service, the service level agreement (SLA) offered by the provider, and the security requirements that the client must meet in order to use the service. When this is done, the service provider adds the client to its access control list. The UXP system provides the technical means to manage the access rights of service clients. Different infrastructures can have different rules for providing access to services.

In the UXP system, members communicate directly without intermediaries. All the messages (requests and responses) are signed and time-stamped and sent over encrypted and mutually authenticated channel.

The governing agency does not take part in the actual message exchange. Instead, it acts as a coordinator and facilitates the communication.

4.1.2 International/Industrial Data Spaces

The Industrial Data Space (IDS) [54] initiative contributes to the design of enterprise architectures in commercial and industrial digitisation scenarios. By providing an architecture for secure exchange of data, the Industrial Data Space bridges the gap between lower-level architectures for communication and basic data services and more abstract architectures for smart data services. It therefore supports the establishment of secure data supply chains from the lowest layer (the data source) to the highest layer (data use), while at the same time ensuring data sovereignty for data owners.

Participation in the Industrial Data Space requires the use of software that is compliant with the Reference Architecture Model. However, the Industrial Data Space is not limited to the software of a specific provider, as it uses an open architecture. This implies that a service can be offered by multiple organisations, including general services in the Industrial Data Space infrastructure, such as a metadata broker or a digital distribution platform.

IDS foresee the following roles (specified from a business perspective):

1. Data Owner

The Data Owner has the legal rights and complete control over its data. Usually, a participant acting as a Data Owner assumes the role of a Data Provider at the same time. However, there may be cases in which the Data Owner is not the Data Provider. The only activity of the Data Owner is to authorise a Data Provider to publish its data. Any authorisation should be documented in a contract, including a policy describing the permissions granted to that specific data. This contract must not necessarily be a paper document but may also be an electronic file.

2. Data Provider

The Data Provider exposes data to be exchanged in the Industrial Data Space. In most cases, the Data Provider is identical with the Data Owner, but not necessarily. To submit metadata to a Broker, or exchange data with a Data Consumer, the Data Provider uses software components that are compliant with the Reference Architecture Model of the Industrial Data Space. Exchanging data with a Data Consumer is the main activity of a Data

Provider. To facilitate a data request from a Data Consumer, the Data Provider must provide metadata about its data to a broker first. However, a Broker is not necessarily required for a Data Consumer and a Data Provider to establish a connection.

3. Data Consumer

The Data Consumer receives data from a Data Provider. The activities performed by the Data Consumer are therefore similar to the activities performed by the Data Provider. Before the connection to a Data Provider can be established, the Data Consumer can search for existing datasets using a Broker. The Broker then provides the required metadata for the Data Consumer to connect to a Data Provider. Alternatively, the connection between the Data Provider and the Data Consumer can be established directly.

4. Broker Service Provider

The main duty of the Broker Service Provider is to manage a metadata repository that provides information about the data sources available in the Industrial Data Space. As the role of the Broker Service Provider is central, but non-exclusive, multiple Broker Service Providers may be around at the same time. The activities of the Broker Service Provider mainly focus on receiving and providing metadata. The Broker Service Provider must provide an interface to receive metadata from the Data Providers.

5. Clearing House

The Clearing House is an intermediary that provides clearing and settlement services for all financial and data exchange transactions. The Clearing House should log all activities performed in the course of a data exchange. After a data exchange has been completed, both the Data Provider and the Data Consumer should confirm transmission and reception of the data, respectively, by logging the transaction at the Clearing House. Based on the logged data, the transaction can be billed then. The logging information can also be used to resolve conflicts (for example, to clarify whether a data package has been received by the Data Consumer or not).

6. Identity Provider

For secure operation, and to avoid unauthorised access to data in the Industrial Data Space, there has to be a service to verify identities. An identity needs to be described by a set of properties (for example, characterising the role of the identity within an organisation). The Identity Provider should offer a service to create, maintain, manage and validate identity information of and for participants in the Industrial Data Space.

7. App Store Provider

The App Store provides applications that can be deployed in the Industrial Data Space to enrich the data processing workflows. An option would be to have the artefacts of an App Store certified by a Certification Body, following the required certification procedures. The App Store Provider is responsible for managing information about Data Apps offered by App Providers. App Providers should describe their Data Apps using metadata, in compliance with a metadata model describing the semantics of the services. The App Store should provide interfaces for publishing and retrieving Data Apps plus corresponding metadata.

8. App Provider

App Providers develop Data Apps to be used in the Industrial Data Space. To be deployable, a Data App has to be compliant with the system architecture of the Industrial Data Space. In addition, Data Apps can be certified by a Certification Body, which would increase the trust in such apps. All Data Apps need to be published in an App Store for being accessed and used by Data Consumers and Data Providers.

9. Vocabulary Provider

The Vocabulary Provider manages and offers vocabularies that can be used to annotate and describe datasets. In particular, the Vocabulary Provider provides the Industrial Data Space Vocabulary. However, other (domain specific) vocabularies can be provided as well.

10. Software Provider

A Software Provider provides software that implements the functionality required by the Industrial Data Space. Unlike Data Apps, software is not provided by the App Store, but delivered and used based on individual agreements between a Software Provider and a software user. The difference between an App Provider and a Software Provider is that App Providers distribute their apps exclusively via the App Store, whereas Software Providers use their usual channels for distribution of their products.

Since IDS is also core component in the Boost 4.0 project more details about how it is mapped to the Boost 4.0 RA are provided in paragraph 6.1.

4.1.3 Linked (meta) data pipeline/Open Micka

Open Micka [55] is a web application for management and discovery geospatial metadata. It is published under BSD license.

It offers the following features:

- OGC Catalogue service (CSW 2.0.2)

- Transactions and harvesting
- Metadata editor
- Multilingual user interface
- ISO AP 1.0 profile
- Feature catalogue (ISO 19110)
- Interactive metadata profiles - management
- WFS/Gazetteer for defining metadata - extent
- GEMET thesaurus built-in client
- INSPIRE registry built-in client
- OpenSearch
- INSPIRE ATOM download service - automatically creation from metadata
- Linked data support
- Additional CSW output formats (HTML, JSON, OAI-PMH, ATOM, KML, GeoDCAT, RDF-XML, RDFa ,RSS channel)
- Additional quarriable
- INSPIRE compliant
- Built-in INSPIRE or user defined - metadata validator
- Other tools (for example service availability)

4.1.4 European Open Science Cloud

European Open Science Cloud (EOSC) [56] provides researchers open services for data storage, management, analysis and reuse across disciplines. EOSC will join existing and emerging horizontal and thematic data infrastructures (for example CERN), bridging today's fragmentation and ad-hoc solutions.

4.1.4.1 Access and interface

Work on the EOSC access and interface has already begun under Horizon 2020 Work Programme 2016-2017: the EOSC-hub project will pilot the common platform and the access to EOSC services; while the *eInfra Central* project provides a first catalogue and access to *eInfrastructure* services. The entry points to the EOSC would be similar but not equivalent, and typically would consist of a web-based user interface, or front-end, which can be tailored to the specific needs and context of particular user communities. In addition, it would comprise a common platform building on the EOSC-hub project, that would be accessible to users via machine-to-machine interfaces and which offers access to shared EOSC resources and to the full range of EOSC services. Services provided under the EOSC would be made accessible via an EOSC portal, based on the work developed by the EOSC-hub and eInfra Central projects. Acting as a universal entry point for all potential users, the portal would have a full-fledged user interface supported by the common platform. A universal entry point usually guarantees that all users have access to the full range of services, irrespective of geographical location or scientific affiliation.

In parallel, established user communities could further develop their existing user interfaces upon the common platform, to offer access to EOSC services within their traditional work environment and give their users a choice as described above.

4.2 Development- Engineering and DevOps

DevOps is a software development methodology that combines software development (Dev) with information technology operations (Ops). The goal of DevOps is to shorten the systems development life cycle while also delivering features, fixes, and updates frequently in close alignment with business objectives [4]. As DevOps is intended to be a cross-functional mode of working, rather than a single DevOps tool, there are sets (or "toolchains") of multiple tools [5]. Such DevOps tools are expected to fit into one or more of these categories, reflective of key aspects of the development and delivery process [6], [7]:

- Code
  Code development and review, source code management tools, code merging.
- Build
  Continuous integration tools, build status.
- Test
  Continuous testing tools that provide feedback on business risks.
- Package
  Artefact repository, application pre-deployment staging.
- Release
  Change management, release approvals, release automation.
- Configure
  Infrastructure configuration and management, infrastructure as code tools.
- Monitor
  Applications performance monitoring, and end-user experience.

Some categories are more essential in a DevOps toolchain than others; especially continuous integration and infrastructure as code [8], [9].

4.2.1 Version Control Systems

Version control systems, also known as revision control or source control systems, are used to track changes to software development projects, and allow team members to change and collaborate on the same files. An indicative list follows:

1. Microsoft Team Foundation Server [57]
2. Git [58]
3. AWS CodeCommit [59]
4. Subversion [60]
5. Helix Core [61]
6. IBM Rational ClearCase [62]

4.2.2 Continuous integration

Continuous Integration (CI) is the practice of continuously integrating the changes made to the project and testing them accordingly at least on a daily basis or more frequently, thus resolving common problems with projects. Several CI tools are available. An indicative list follows:

1. Jenkins [63]
2. TeamCity [64]
3. Travis CI [65]
4. Go Continuous Delivery [66]
5. GitLab CI [67]

4.2.3 Continuous testing

Continuous Testing is the process of executing automated tests as part of the software delivery pipeline in order to obtain feedback on the business risks associated with a software release candidate as rapidly as possible. It evolves and extends test automation to address the increased complexity and pace of modern application development and delivery. An indicative list of tools is as follows:

1. Selenium With Robot Framework [68]
2. Testsigma [69]
3. IBM Rational Functional Tester [70]
4. Tricentis Tosca [71]
5. Unified Functional Test (UFT) [72]

4.2.4 Artefact repositories

An artefact repository is a collection of binary software artefacts and metadata stored in a defined directory structure which is used by clients such Maven, Mercury, or Ivy to retrieve binaries during a build process. An indicative list of artefact repositories is as follows:

1. JFrog Artifactory [73]
2. Sonatype Nexus [74]

3. Apache Archiva [75]

4.2.5 Application release automation

Application-release automation refers to the process of packaging and deploying an application or update of an application from development, across various environments, and ultimately to production [10]. Such solutions must combine the capabilities of deployment automation, environment management and modelling, and release coordination [11]. Application release automation tools are useful in accelerating application deployments, and provide value for DevOps teams by addressing the ability to:

- The ability to plan, schedule, and track application code releases
- Application deployment automation
- Integration with existing tools and infrastructure
- The ability to scale across the enterprise

An indicative list of tools is as follows:

1. XL Deploy & XL Release [76]
2. Visual Studio Release Management [77]
3. IBM UrbanCode Release Automation [78]
4. CA Release Automation and Automic [79]
5. ElectricFlow [80]

4.2.6 Configuration management

Configuration management is a governance and systems engineering process for ensuring consistency among physical and logical assets in an operational environment. The configuration management process seeks to identify and track individual configuration items, documenting functional capabilities and interdependencies. Administrators, technicians and software developers can use configuration management tools to verify the effect a change to one configuration item has on other systems. An indicative list of configuration management tools is as follows:

1. Ansible [81]
2. Chef [82]
3. Puppet [83]
4. SaltStack [84]

4.2.7 Application Performance Monitoring

Application performance management (APM) is the monitoring and management of performance and availability of software applications. APM strives to detect and diagnose complex application performance problems to maintain an expected level of service. An indicative list of APM tools is as follows:

1. New Relic [85]
2. AppDynamics [86]
3. Compuware APM [87]

4.3 Communication

Communication technologies and protocols are essential for the Industry 4.0 concept, since the blending of the digital world with the physical through Internet-based technologies is at the heart of the intelligent factory era. Flexibility, efficient use of resources, improved ergonomics, and integrating customers and business partners into the business and value creation processes, are all features of this blending process. In fact, the Industry 4.0 includes the amalgamation of Cyber-Physical Systems (CPS), Internet of Things (IoT), Internet of Services (IoS), and cloud computing, and the ways these work together. Industry 4.0 also comprises the notion of the modular-structured smart factory, where cyber-physical systems monitor physical processes, create a virtual copy of the physical world, and make decentralised decisions.

Information and communication technologies play a key role in implementing these Industry 4.0 concepts. Cyber-physical systems such as sensors, actuators, embedded computers, smartphones and machines are connected to one another and exchange data both between themselves and outside of factories' shop floors. Today's automation networks and fieldbus systems must therefore not only guarantee that machines and facilities can carry out production with safety, precision and efficiency, but they must also help towards establishing a universal solution for integrating different IT systems on different rings of the organisational ladder within a factory, company or business.

This Section will give an overview of existing communication models and technologies, while also describing the communication challenges and barriers for IoT and Industry 4.0 concepts, as well as give some recommendations on how to mitigate some of the presented issues.

4.3.1 Overview of Communication Models

Several communication models are in use today in factories all around the world. Communication is central to both IoT and Industry 4.0. Networking technologies enable devices in the shop floor to communicate with other devices as well as with applications and services that are running in the cloud, whether on factory, company or business levels. The internet relies on standardised protocols to ensure that communication between heterogeneous devices can occur securely and reliably. Standard protocols specify the rules and formats that devices use for establishing and managing networks, as well as for transmission of data across those networks.

4.3.1.1 Models Overview

The Open Systems Interconnection (OSI) model is an ISO-standard abstract model that defines a networking framework to implement protocols in layers, with control passed from one layer to the next, describing it as a stack of seven protocol layers. From the top down, these layers are: application, presentation, session, transport, network, data link and physical. The OSI model was originally conceived as a standard architecture for building network systems and indeed, many popular network technologies today reflect the layered design of OSI.

While the OSI model is quite an abstract and academic concept, rarely encountered outside books and articles about computer networking, the TCP/IP model is a simpler, easier-to-understand, and more practical proposition: it is the bedrock of the Internet. This model provides end-to-end data communication specifying how data should be packetized, addressed, transmitted, routed, and received. This functionality is organised into four abstraction layers, which classify all related protocols according to the scope of networking involved.

4.3.1.2 Mapping the different models to IoT Protocols and Technologies

Many emerging and competing networking technologies are being adopted within the IoT space. Multiple technologies are offered by different vendors or are aimed at different vertical markets like home automation, healthcare, or industrial IoT, often provide alternative implementations of the same standard protocols. Technologies used for internet connectivity, like Ethernet, for example, can often be applied within the IoT; however, new technologies are being developed specifically to meet the challenges of IoT.

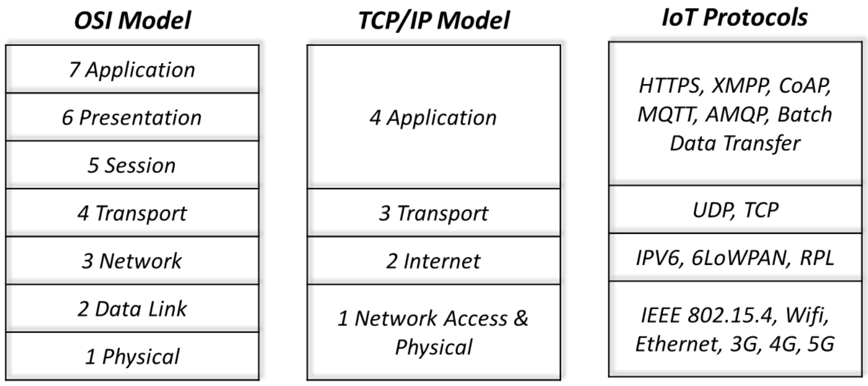

Figure 3. Mapping different communication models to IoT Protocols

Further down the stack towards physical transmission technologies, companies face more challenges that are specific to IoT devices and IoT contexts. IoT network technologies to be aware of toward the bottom of the protocol stack include cellular, Wi-Fi, and Ethernet, as well as more specialized solutions such as LPWAN, Bluetooth Low Energy (BLE), ZigBee, NFC, and RFID. Table 1 presents some examples of technologies per TCP/IP layer.

Layer	Technologies
Network Access & Physical Layer	Bluetooth (IEEE 802.15.1), BLE
	Near field communication (NFC)
	IEEE 802.15.4, ZigBee, WirelessHART
	Ethernet (IEEE 802.3)
	Wi-Fi (IEEE 802.11)
	3G, 4G, 5G
Internet Layer	IPv4, IPv6
	CoAP
	RPL
Application Layer	HTTP and HTTPS
	CoAP, 6LoWPAN over UDP
	MQTT
	AMQP
	XMPP
	Apache Kafka
	Apache Flume
	Apache Sqoop

Table 1 Communication Technologies per TCP/IP Layer

4.3.2 Communication Considerations, Challenges and Recommendations

The choice and usage of a network depend on range, bandwidth, power usage, connectivity resilience, interoperability and security. To improve range, networks should be designed to get the data from devices at shop floor level to where it will be used, whether at factory, company, business or market level. An important consideration is selecting the appropriate network protocol to match the range requirements. Using edge computing, which moves the analysis to the data out of the devices rather than moving the data itself elsewhere for processing, can improve range.

Bandwidth, or the amount of data that can be transmitted in a specific period of time, limits the rate at which data can be collected from IoT devices or other data producers and transmitted upstream. Several factors should be considered: (i) the volume of data that each device or producer is generating, (ii) the number of devices that are deployed in a network, and (iii) whether the data is being sent as a constant stream or in intermittent bursts, as the bandwidth that is available will need to cope with the peak periods.

The packet size of the chosen networking protocol should match with the size of the data that is being transmitted. It is inefficient to send packets padded out with empty data, but on the other side there are overheads in splitting larger chunks of data up across too many small packets. Additionally, data transmission rates are not always symmetrical (upload rates might be slower than download rates). So, if two-way communication between devices is employed, data transmission needs to be factored in. Wireless and cellular networks are traditionally low-bandwidth, so one should consider whether a wireless technology is the right choice for high-volume applications.

Moreover, another important consideration has to do with the amount of the raw data that needs to be transmitted. One solution might be to capture less data by sampling less frequently, capturing fewer variables, or performing some filtering on the device to remove insignificant data. If the data is aggregated before transmission, it helps reduce the volume of data to be transmitted, but this process might have implications on flexibility and granularity in the upstream analysis. Data aggregation and bursting is not always suitable for time-sensitive or latency-sensitive data either. All of these techniques also increase the data processing and storage requirements for the device.

Transmitting data from a device consumes power, and transmitting data over long ranges requires more power than over a short range. One must consider the devices that operate on battery power to conserve power and prolong the life of the battery thus reducing operating costs. To prolong battery life, the device can be put into sleep mode whenever it is idle. A good practice is to model the energy consumption of the device under different loads and different network conditions to ensure that the device's power supply and storage capacity matches with the power that is required to transmit the necessary data by using the networking technologies that were adopted. IoT devices are not always connected. In some cases, devices will connect periodically by design in order to save power or bandwidth. However, sometimes an unreliable network might cause devices to drop off due to connectivity issues. Sometimes quality of service issues, such as dealing with interference or channel contention on a wireless network using a shared spectrum.

With so many different devices connecting to the IoT, interoperability can be a challenge. Adopting standard protocols has been the traditional approach for maintaining interoperability on the internet. However, for the IoT, standardisation processes sometimes struggle to keep up with the rapid pace of change and technologies are released based on upcoming versions of standards that are still subject to change. In these cases, one should consider the ecosystem around the technologies; that is, ask these questions: Are they widely adopted? Are they open versus proprietary? How many implementations are available?

Finally, security is always a priority, so the selection of the appropriate networking technologies that implement end-to-end security, including authentication, encryption, and open port protection is recommended. For example, IEEE 802.15.4 includes a security model that provides security features that include access control, message integrity, message confidentiality, and replay protection, which are implemented by technologies based on this standard such as ZigBee. In terms of authentication, adopt secure protocols to support authentication at the device level, for gateways, users, and applications and services. For example, consider adopting the X.509 standard for device authentication.

If using Wi-Fi, a network can use Wireless Protected Access 2 (WPA2) for wireless network encryption or might adopt a Private Pre-Shared Key (PPSK) approach. To ensure privacy and data integrity for communication between applications, be sure to adopt TLS or Datagram Transport-Layer Security (DTLS), which is based on TLS, but adapted for unreliable connections that run over UDP. TLS encrypts application data and ensures its integrity. Port protection ensures that only the ports that are required for communication with the gateway or upstream applications or services remain open to external connections. All other ports should be disabled or protected by firewalls. For example, device ports might be exposed when exploiting Universal Plug and Play (UPnP) vulnerabilities, so UPnP should be disabled on the router.

4.4 Standards

Smart sensors, robotics, AI, big data lakes and cloud computing are helping to pave the way for gains in productivity, financial and operational performance, output, and market share as well as improved control and visibility throughout the supply chain.

Manufacturing agility at supplying highly customized products depends on rich information systems that control every aspect of operations, from product design, customer orders, supply chain management, just in time and just in sequence assembly, production cells with symbiotic robots and human workers, distribution and post sales for servicing, customer relationship management and product improvements, through software and hardware upgrades.

Industry 4.0 builds upon ideas by Michael Porter on the role of value chains for competitive advantage [119], emphasizing the need for horizontal integration across organizations in the supply chain, organizations in the enterprise itself, and those involved in the distribution and post sales processes. Vertical integration is needed to bridge the different systems used from the shop floor to the office floor (see Figure 4).

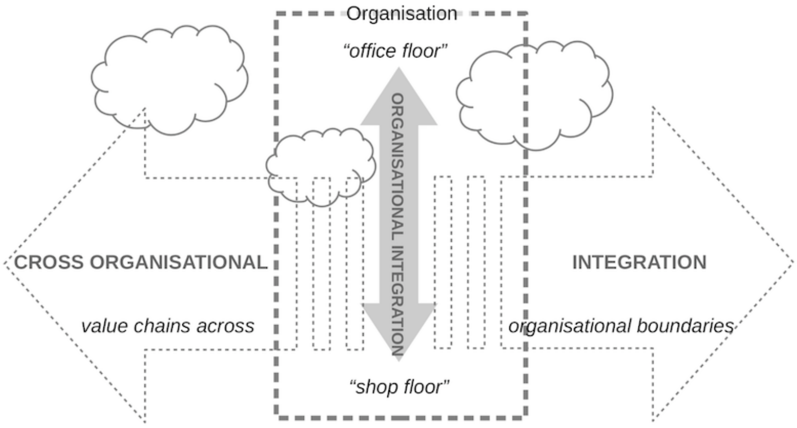

Figure 4 Horizontal and Vertical Integration

This level of integration is challenging to achieve in a way that provides agility in respect to continuously evolving requirements. Traditional technologies and methodologies are inadequate, and this is driving interest in graph data as compared to tabular data used with relational database management systems. Graph data is better suited for combining information from heterogeneous data silos with ever changing requirements. This becomes feasible by adopting an enterprise wide approach to data management and governance. This in turn relies on developing control over metadata and abstracting from the complexity inherent to data silos and the formats they use. Knowledge Graph is a popular term for expressing this metadata as a graph of nodes and links.

- Knowledge Graphs represent a collection of interlinked descriptions of entities – real-world objects, events, situations or abstract concepts, **Ontotext**;
- Knowledge Graphs are the only realistic way to manage enterprise data in full generality, at scale, in a world where connectedness is everything, **Kendall Clark**, 26 Jun 2017

Enterprise knowledge graphs can be built using W3C's approach to graph data (RDF/Linked Data). Information services can be simplified through abstractions that hide the underlying protocols and data formats – W3C's Web of Things.

The need for enterprise wide oversight and management of data has given rise to the field of Data Governance. The Data Management Association provides extensive guidance, including the following diagram (Figure 5) which lists the many aspects involved:

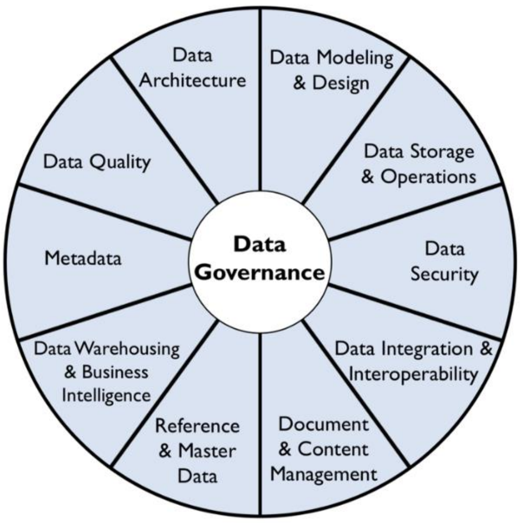

Figure 5 DAMA Wheel

The European Commission launched the Digitising European Industry (DEI) initiative in April 2016 with the aim of boosting competitiveness through adoption of advanced digital technologies throughout the enterprise and enabling the Digital Single Market across EU member states. Digitisation is not new, but hitherto has been isolated into separate business functions rather than forming an integrated whole. Examples include Enterprise Resource Planning (ERP) and Manufacturing Execution Systems (MES), critical planning, scheduling, warehousing, inventory management, and logistics processes, which have been automated and simplified through the use of digitisation.

The European Rolling Plan on ICT standardisation is being maintained with the help of the European Multi-Stakeholder Platform (MSP). The plan lists topics identified as EU policy priorities where standardisation, standards or ICT technical specifications ought to play a key role in the implementation of that policy. It also contains a list of ongoing or notable actions in the areas of interest, including DEI. Of note is the June 2018 workshop in which the European Commission presented the interim results of the MSP-DEI Working Group. The MSP-DEI WG is working on:

1. Identifying standardisation needs for manufacturing sector
2. Landscaping ongoing standardisation activities, fora & consortia, Large Scale Pilots, Public-Private Partnerships (PPPs), DE/IT/FR trilateral cooperation, and other research projects, etc. that are relevant to the digitalisation of European industry
3. Developing a model for synchronisation of standardisation activities
4. Proposing a roadmap, taking national standardisation roadmaps into account and specifying concrete actions for inclusion in the Rolling Plan on ICT standardisation

The standardisation ecosystem is split into formally recognised *(de jure)* standards bodies such as IEC, ISO, ITU, CENELEC, CEN, ETSI, 3GPP and oneM2M; and *(de facto)* standards bodies that play a key role, but which are not formally recognised, such as W3C, OASIS, OMG, IETF, IEEE and the OPC Foundation. This distinction is made by WTO, EU Regulation 1025/2012, national government contracts and rules. Other distinctions include:

- The distinction between standards for meeting regulatory requirements, and standards defining technical specifications that enable interoperability
- The difference between horizontal standards and technical specifications
- Between normative standards and technical specifications that define the basis for compliance and informative publications which may serve many different purposes

Figure 6 illustrates the different categories of standards considered by the DEI-MSP WG.

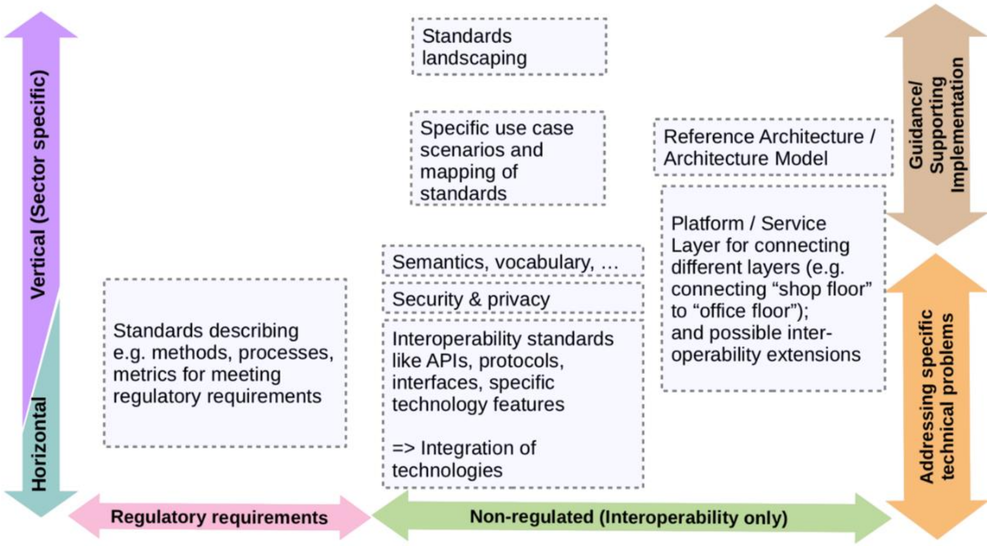

Figure 6 Standardisation and DEI – Jochen Friedrich, DEI-MSP WG, June 2018

The final report of the MSP DEI Working Group was issued end of November 2018. Task 1 & 2 of the Working Group was to create Recommendations around landscape and gaps in standardisation. They gave a good overview of the landscape of current initiatives including a list of Specification Developing Organisations (SDOs), Reference models, Standardisation initiatives such as oneM2M and political initiatives like the trilateral activities around the topic in Germany, France and Italy. Both tasks ended with the recommendation of nine actions:

1. Common communications standards and a reference architecture for connections between machines (M2M) and with sensors and actuators in a supply chain environment are a basic need and a priority.

2. Check whether the Skills – Agenda of the EC contains the appropriate topics and standards

3. Conduct a study to identify and analyse opportunities for revisions of existing standards

4. Improve interoperability and reduce overlap, redundancy and fragmentation.

5. Interoperable and integrated security - SDOs should work on interoperability standards for security and for linking communication protocols in order to provide end-to-end security for complex manufacturing systems including the span of virtual actors (from devices and sensors to enterprise systems).

6. Create a hierarchical catalogue of technical and social measures for assuring privacy protection and task all SDOs impacting the DEI domain in general and the

advanced manufacturing domain in particular to comment on and prioritize the elements in the catalogue.

7. Standards should be developed to define the main characteristics for all levels of the interaction from mechanical to electrical to protocol to semantic levels between robot and tool to ensure the exchangeability and to enable the design of generic tooling (plug-and-play).

8. Start the discussion about the possible development of harmonised standards in the area of additive manufacturing.

9. Develop standards for ensuring long-term traceability of material to enable re-use and recycling.

Many of those suggestions concern metadata and align well with the IDS model used by Boost 4.0. There is insufficient space here to delve into the details of the different standards organisations and technologies. For that you are recommended to read Boost 4.0 deliverable D2.7 "standardisation and certification".

4.5 Cybersecurity and Trust

Exchange, storage and processing of large amounts of data with both Cloud and Big Data Environments can expose the data to security vulnerabilities, such as:

- Weak management of access, credentials and identification.
- Unsafe APIs.
- Vulnerabilities of systems, communications and applications.
- Accounts hacking.
- Social Engineering (Phishing, Baiting …)
- Malware and viruses (Ransomware, Trojans …).
- Advanced Persistent Threats (APT).
- Data Loss.
- Abuse and harmful use of cloud services.
- Denial of Service (DoS), Distributed Denial of Services (DDoS).
- Shared technology problems.

4.5.1 Mechanisms of Prevention, Detection and Action against Cyber-attacks in Cloud and Big Data Environments

The following mechanism may address the vulnerabilities outlined above:

- **Anti-phishing.** Most of these programs work by identifying phishing content on websites and emails; Some anti-phishing software can for example, integrate with web browsers and email clients as a toolbar that shows the real domain of the visited site.
- **Antivirus.** Antiviruses are programs whose objective is to prevent, detect or eliminate computer viruses that may affect the equipment at an End-Point level.
- **Unified Threat Management (UTM) [21].** It is a term that refers to a single network device with multiple functions. The basic functionalities that it must have are: Antivirus, Firewall and intrusion detection and prevention systems.
- **Firewall [22].** This type of cybersecurity solution allows to manage and filter all incoming and outgoing traffic (connections) between a given network and the internet.
- **Intrusion Detection System [23] /Intrusion Prevention Systems.** They increase the security of the networks, monitoring traffic, examining and analysing packets for suspicious data. Both systems base their detections mainly on signatures already detected and recognised. In fact, there are some repositories where these signatures are stored and added.
    - The Intrusion Detection System provides the network with a degree of security of a preventive nature in the face of any suspicious activity. However, unlike the IPS system, it is not designed to stop or block these attacks.
    - The Intrusion Prevention System is designed to analyse the attack data and act accordingly, stopping it at the same time it is being gestated and before it succeeds.
- **SIEM (Security Information and Event Management) [24].** Most SIEM systems work by deploying multiple collection agents in a hierarchical manner to collect events related to the security of end-user devices, servers, network equipment and even specialised security equipment such as firewalls, antivirus or intrusion prevention systems. Collectors send events to a centralised management console, which performs inspections and detect anomalies. Most SIEMs work with a series of previously defined or that can be defined rules (some SIEMs provide a UI for the introduction of new rules manually) taking into consideration the infrastructure and the activity of a given company. However, currently also stand out SIEMs that work autonomously and are able to detect threats, anomalies or even Malware automatically by using Machine Learning and Deep Learning, and monitoring the computational performance of the equipment connected to the network for instance.

- **Information Encryption techniques and tools.** These encryption tools and techniques are particularly conducive to the scenario that is intended to be addressed within the BOOST4.0 project, where it is intended to operate with large volumes of data, and it is necessary to guarantee the confidentiality, availability and integrity of this data. In this way, there are different types of techniques for data and information encryption.
    - **Symmetric Encryption [25].** Symmetric encryption, also known as "Shared Key" or "Shared Secret", is one in which a single key is used to encrypt and decrypt information. Among the symmetric encryption algorithms *DES [26]*, *3DES [27]*, *AES [28]* and *RC4 [29]* are mentioned. The 3DES and AES algorithms are commonly used by the IPSEC [30] protocol to establish VPN connections. The RC4 algorithm is used in wireless networking technologies for the encryption of information in the WEP and WPA protocols.
    - **Asymmetric Encryption [31].** Asymmetric encryption, also known as "Public Key Cryptography", differs mainly from its counterpart in symmetric encryption, in that it uses two keys instead of one to exercise the functions of encryption and decryption of information. In the asymmetric encryption, each user who intervenes in the communication counts with two keys: a public key and a private key. The public key is used to encrypt the information and private key to decrypt the information. Some of the best known and most used algorithms for asymmetric encryption are the *RSA [32]* or *Diffie-Hellman [33]*. The biggest advantage of asymmetric cryptography is that the distribution of keys is easier and safer since the key that is distributed is the public key, keeping the private key for the exclusive use of the owner
    - **Homomorphic Encryption [34].** The homomorphic encryption has two forms, partially homomorphic encryption and fully homomorphic encryption. The partially homomorphic encryption is defined when it has limits on the amount of transactions with encrypted data. However, the fully homomorphic encryption is a cryptographic system that allows a set of arbitrary mathematical operations (without limitation) to be made in the *ciphertext [35]*, which should result in another *ciphertext* corresponding to the result of the operation in plain text.
- **Authentication, Authorisation and Access control mechanisms.** In the current market there are several authentication mechanisms such as Single Factor Authentication, two-factor authentication or multi-factor authentication. The *OpenID [36]* protocol is generally the more used protocol for solutions that offer authentication capabilities. On the other hand, as far as authorisation is concerned,

the most well-known and used protocol is called *OAuth [37]* and allows, among other things, to apply and define different types of permits between two different services or applications.

- **Virtual Private Networks (VPN) [39].** It allows the computer in the network to send and receive data over shared or public networks as if it were a private network with all the functionality, security and management policies of a private network. Moreover, these VPNs have some additional security features, such as authentication and authorisation, to make sure that users at the other end are authorised to use this network and what level of access they have. On the other hand, algorithms such as *Message Digest* (MD5) [40] and *Secure Hash Algorithm* (SHA) [41] are also used to guarantee the integrity of the data that is sent, preventing this data from being altered. Finally, it also uses encryption mechanisms such as *DES* or *3DES* to guarantee that the information circulating through these VPNs can only be interpreted by the recipient of the same, thus guaranteeing the confidentiality/privacy of this information. These VPN environments may be suitable for Big Data environments in which it is necessary to connect remotely to a network, in order to collect certain information, or for environments where information exchange of a sensitive nature is periodically and continuously carried out between two specific networks, for example between two headquarters of the same company.
- **Network traffic monitoring and reporting tools.** They are based on the collection and analysis of network traffic in order to obtain information. For this purpose, various devices are used, such as sniffers, routers, computers with traffic analysis software or devices with support for *SNMP [42]*, *RMON [43]* or *Netflow [44]*. Unlike active monitoring techniques, passives do not add traffic to the network. Its use is usually focused on the characterisation of traffic on the network and to account for its use. Besides that, it also aims to detect possible threats or anomalies within a network, guaranteeing communications through it. On the other hand, it also offers metrics that are of interest from the point of view of the network performance. Therefore, these types of solutions are considered essential in networks where there is a large flow of information, and where the information exchanged is in some way sensitive. That is why it can be considered as essential tool for monitoring the infrastructure that is planned to be deployed and adopted within the BOOST4.0 project.

5 Alignment with existing Reference Architecture and models

To achieve its purpose to support the big data lighthouse smart connected factory pilots, Boost 4.0 RA has to be aligned with existing reference architectures and models. This section outlines how the Boost 4.0 RA is aligned with the Big Data Value Association (BDVA) Big Data Value Reference Model, the Reference Architectural Model Industrie 4.0 (RAMI4.0), and the Industrial Internet Reference Architecture (IIRA).

5.1 BDVA Big Data Value Reference Model

The BDVA Big Data Value Reference Model is illustrated in below (Figure 7) and detailed in the BDVA SRIA [12]:

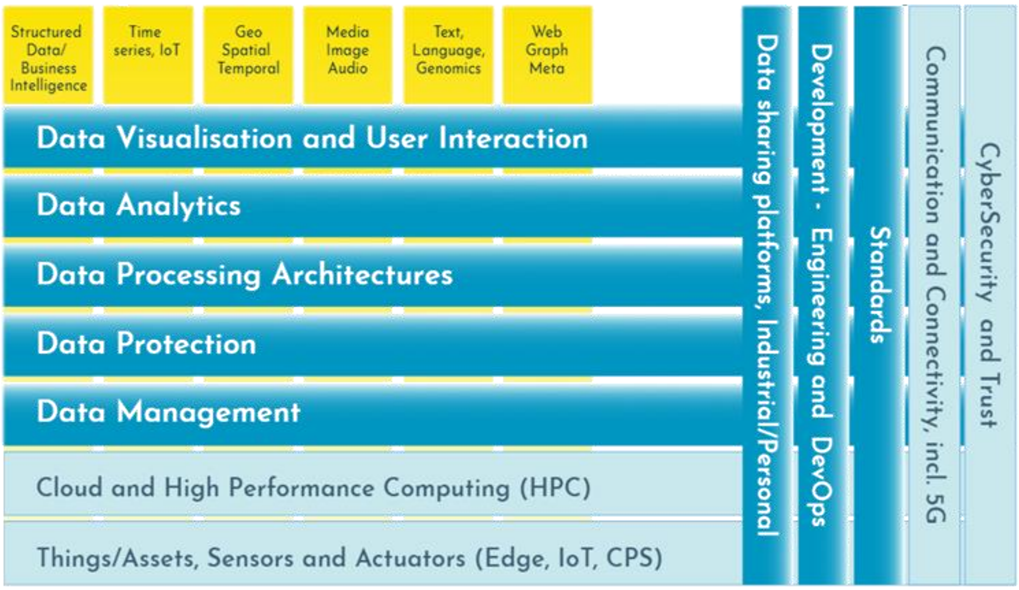

Figure 7 BDVA Big Data Value Reference Model

The model consists of the following horizontal concerns:

1. *Data Visualisation and User Interaction*

Advanced visualisation approaches for improved user experience.

2. *Data Analytics*

Data analytics to improve data understanding, deep learning and the meaningfulness of data.

3. *Data Processing Architectures*

Optimised and scalable architectures for analytics of both data-at-rest and data-in-motion, with low latency delivering real-time analytics.

4. *Data Protection*

Privacy and anonymization mechanisms to facilitate data protection. This is shown related to data management and processing as there is a strong link here, but it can also be associated with the area of cybersecurity.

5. *Data Management*

Principles and techniques for data management.

6. *The Cloud and High Performance Computing (HPC)*

Effective Big Data processing and data management might imply the effective usage of Cloud and High Performance computing infrastructures.

7. *IoT, CPS, Edge and Fog Computing*

A main source of Big Data is sensor data from an IoT context and actuator interaction in Cyber Physical Systems. In order to meet real-time needs, it will often be necessary to handle Big Data aspects at the edge of the system.

BDVA also consist of vertical concerns that cover all horizontal concerns, these are:

1. *Big Data Types and Semantics*

The following 6 Big Data types have been identified, based on the fact that they often lead to the use of different techniques and mechanisms in the horizontal concerns, which should be considered, for instance, for data analytics and data storage: (1) Structured data; (2) Time series data; (3) Geospatial data; (4) Media, Image, Video and Audio data; (5) Text data, including Natural Language Processing data and Genomics representations; and (6) Graph data, Network/Web data and Metadata. In addition, it is important to support both the syntactical and semantic aspects of data for all Big Data types.

2. *Standards*

Standardisation of Big Data technology areas to facilitate data integration, sharing and interoperability.

3. *Communication* and Connectivity

Effective communication and connectivity mechanisms are necessary in providing support for Big Data.

4. *Cybersecurity*

Big Data often need support to maintain security and trust beyond privacy and anonymization. The aspect of trust frequently has links to trust mechanisms such as Blockchain technologies, smart contracts and various forms of encryption.

5. Engineering and DevOps for building Big Data Value systems

Big Data technologies have gained significant momentum in research and innovation. However, mature, proven and empirically sound engineering methodologies for building next-generation Big Data Value systems are not yet available. Also, there is a lack of proven approaches for continuous development and operations (DevOps) of Big Data Value systems. The availability of engineering methodologies and DevOps approaches – combined with adequate tool chains and Big Data platforms – will be essential for fostering productivity and quality.

6. Marketplaces, Industrial Data Platforms and Personal Data Platforms (IDPs/PDPs),

Ecosystems for Data Sharing and Innovation Support

Data platforms for data sharing include, in particular, IDPs and PDPs, but also other data sharing platforms like Research Data Platforms (RDPs) and Urban/City Data Platforms (UDPs). These platforms facilitate the efficient usage of a number of the horizontal and vertical Big Data areas, most notably data management, data processing, data protection and cybersecurity.

The following tables detail how the Big Data Value Reference Model concerns are addressed in Boost 4.0 RA:

Table 2 Big Data Value Reference Model horizontal concerns alignment

Big Data Value Reference Model	Boost 4.0 RA
The Cloud and High Performance Computing (HPC), IoT, CPS, Edge and Fog Computing	**External Data Sources, Infrastructure** These elements represent the actual systems that Boost 4.0 platform has to interface with in order to access information. It also defines the infrastructure required for providing added value services.
-	**Integration Layer** Covers mechanisms access & control of external data sources and infrastructure.
Data management	**Information and Core Big Data Layers**: Data management The sub-layer addresses the data management needs of the architecture.
Data Protection	Mechanisms for data protection are covered in the cross-cutting aspect Cybersecurity and Trust

Data processing	**Information and Core Big Data Layers**: Data management The sub-layer addresses the data management needs of data processing.
Data Analytics	**Information and Core Big Data Layers**: Data analytics The sub-layer addresses the data management needs of data analysis.
Data visualisation and user interaction	**Information and Core Big Data Layers**: Data visualisation The sub-layer addresses the mechanisms required to handle big data visualisation.

Table 3 Big Data Value Reference Model vertical concerns alignment

Big Data Value Reference Model	Boost 4.0 RA
Standards	**Standards** Aligns standards use throughout the different layers.
Communication and Connectivity	**Communication** Defines technologies and APIs to be used on the different layers of the architectures.
Cybersecurity	**Cybersecurity & Trust** Covers topics such as device and application registration, identity and access management, data governance, data protection, and so on, and is responsible for synchronising theses mechanisms throughout the layers.
Engineering and DevOps for building Big Data Value systems	**Development - Engineering and DevOps** Covers tool chains and frameworks that significantly increase productivity in terms of developing and deploying big data solutions. It also includes testing, monitoring and verification tools and methodologies to increase reliability, security, energy efficiency and quality of the developed solutions.
Marketplaces, Industrial Data Platforms and Personal Data Platforms	**Data Sharing Platforms** This aspect considers the needs for allowing data providers to share their data as a commodity, covering specific data, for a predefined space of time, and with a guarantee of reversibility at the end of the contract.

5.2NIST Big Data Reference Architecture

The NBDRA [13] consists of five main components, shown in Figure 8, that represent different technical roles that exist in every Big Data system. These functional components are:

- System Orchestrator: Defines and integrates the required data application activities into an operational vertical system.
- Data Provider: Introduces new data or information feeds into the Big Data system.
- Big Data Application Provider: Executes a data life cycle to meet security and privacy requirements as well as System Orchestrator-defined requirements.
- Big Data Framework Provider: Establishes a computing framework in which to execute certain transformation applications while protecting the privacy and integrity of data;
- Data Consumer: Includes end users or other systems that use the results of the Big Data

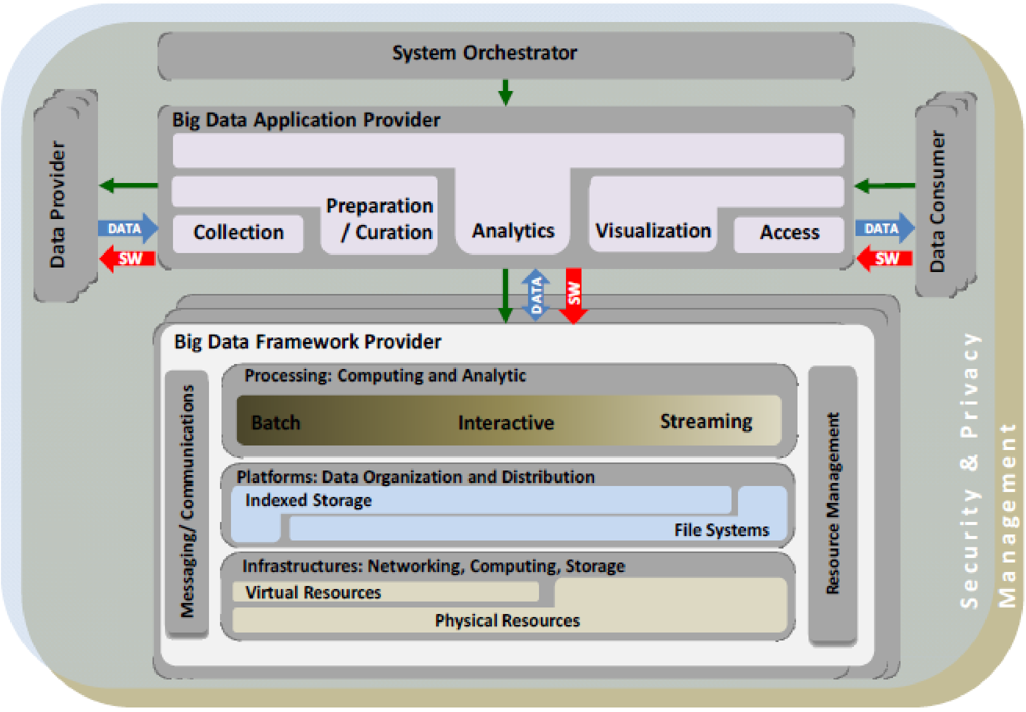

Figure 8 NBDRA conceptual model

The two fabrics shown in Figure 8 encompassing the five functional components are:

- Management
- Security and Privacy

These two fabrics provide services and functionality to the five functional components in the areas specific to Big Data and are crucial to any Big Data solution.

The following table details the mapping between Boost 4.0 RA and NBDRA:

Table 4 NBDRA alignment

NBDRA	Boost 4.0 RA
Data Provider	External Data Sources The external data sources element acts as the source of big data in the Boost 4.0 RA. These elements are further specialised into entities associated with the manufacturing domain (for example sensors, MES, ERP, and so on). This element corresponds directly to the NBDRA Data Provider.
Infrastructures	Infrastructure This element provides the resources necessary to realise the functions of the components of the all layers above. This element corresponds directly to the NBDRA Infrastructures.
Processing: Computing and analytics	Data processing architectures layer This layer groups together components to support implementation of applications that can deal with the volume, velocity, variety, and variability of data. Therefore, it corresponds directly to the NBDRA Processing: Computing and analytics.
Collection, preparation/ curation	Data Management layer This layer groups together components facilitating data collection, preparation, curation, linking, and sharing, thus supporting the activities of data collection, preparation, and curation.
Analytics	Data Analytics layer This layer groups together components to support data analysis and thus covers the Analytics activities of NBDRA.
Visualisation	Data visualisation layer This layer groups together the components that support big data visualisation and therefore address the NBDRA activities of visualisation.
Access	Application layer The application layer is responsible for exposing the functionality of the layers below towards potential data consumers. This layer covers the NBDRA Access activities.
System Orchestrator	Development, Engineering, and DevOps crosscutting aspect

	This aspect covers management and orchestration of the whole big data system.
Management, Security and Privacy	**Development, Engineering, and DevOps, Cybersecurity and Trust crosscutting aspects** These aspects are responsible for addressing management, security and privacy considerations.

5.3 RAMI4.0

RAMI 4.0 [14] is constructed from three dimensions such as Layers, Life Cycle & Value Steam and Hierarchy Levels, as shown in Figure 9.

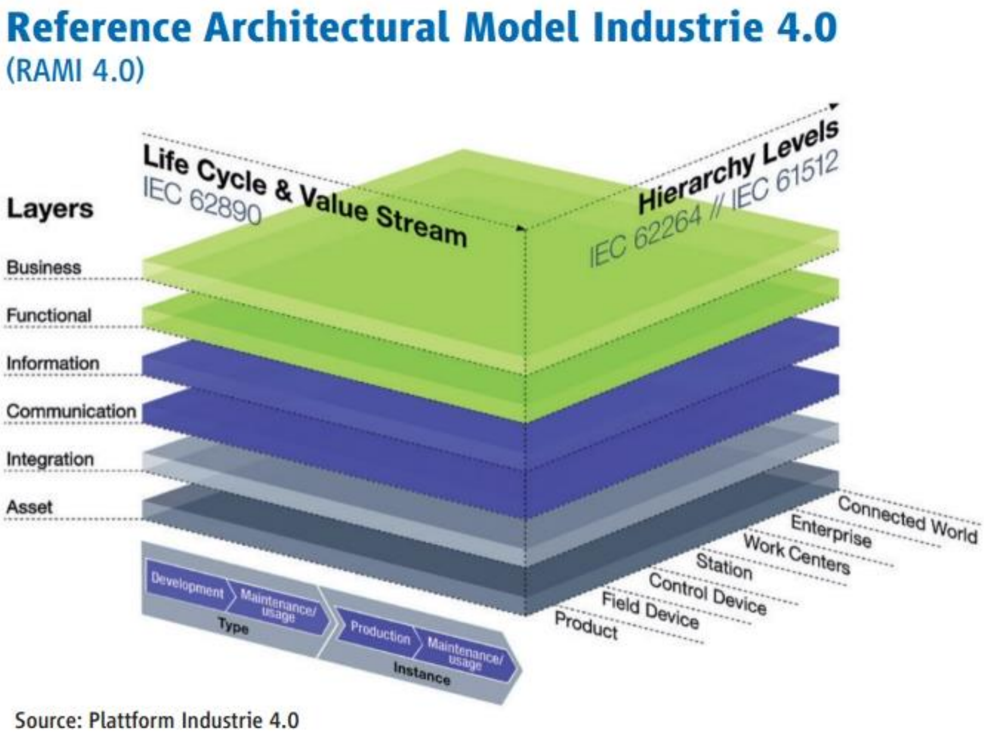

Figure 9 RAMI4.0

5.3.1 Layers

Layers are used in the vertical axis to represent the various perspectives, such as data maps, functional descriptions, communications behaviour, hardware/ assets or business processes. This corresponds to IT thinking where complex projects are split up into clusters of manageable parts.

Business Layer

This layer has the following functionality:

- Ensures the integrity of the functions in the value stream.
- Maps out the business models and the resulting overall process.
- Provides legal and regulatory framework conditions and models of the rules which the system has to follow.
- Orchestrates services in the Functional Layer.
- Provides a link between different business processes.
- Receives events for advancing of the business processes.

The Business Layer does not concern concrete systems, such as an ERP system. ERP functions in the process context are typically located in the Functional Layer.

The business layer of the Boost 4.0 RA combines the functionalities of the layers below to realise business processes and links different business processes in support of the business models of each pilot. Therefore, the Boost.4.0 RA Business Layer corresponds to the RAMI4.0 Business Layer.

Functional Layer

This layer performs the following:

1. Horizontal integration of the various functions.
2. Run time and modelling environment for services which support business processes.
3. Run time environment for applications and technical functionality.
4. Execution of rules and decision-making logic.

The Asset Layer and Integration Layer may also be accessed temporarily for maintenance purposes. Such access is in particular used to call up information and processes which are only relevant to subordinate layers. Examples include flashing of sensors/actuators or the reading of diagnosis data.

The application layer of the Boost4.0 RA groups together components providing pilot case specific logic, and logic supporting the business processes. Additionally, it includes components that facilitate user interaction (input) and services. Its purpose therefore fits in with the Functional layer of RAMI4.0.

Information Layer

This layer provides:

1. Run time environment for (pre-) processing of events.

2. Execution of event-related rules along with their formal descriptions.
3. Event pre-processing for content extraction (generate more complex events or higher quality data).
4. Receiving events and transforming them to match the data which are available for the Functional Layer.

Information representation, processing, transformation, and storage is handled in the Information and Core Big Data Layers of the Boost4.0 RA therefore it directly corresponds to the RAMI4.0 Information layer. It should be noted that because Boost 4.0 RA focuses on big data this layer is further specialised in sub-layers dealing with different aspects of managing/handling big data.

Communication Layer

Standardisation of communication, using a uniform data format, in the direction of the Information Layer. Provision of services for control of the Integration Layer.

The communication cross-cutting aspect of the Boost4.0 RA d defines the technologies and APIs to be used on the different layers of the architectures. Some parts of the RAMI4.0 communication layer are handled in the Communication cross-cutting aspect of Boost 4.0 RA. However, the Communication cross-cutting aspect also affects the rest of the layers in Boost 4.0 and not only the integration and information layers.

Integration Layer

This has the following functionality:

1. Provision of information on the assets (physical components / hardware / documents / software, and so on) in a form which can be processed by computer.
2. Computer-aided control of the technically process.
3. Generation of events from the assets.
4. Contains the elements connected with IT, such as RFID readers, sensors, HMI, and so on Interaction with humans also takes place on this level, for instance via the Human Machine Interface (HMI).

This layer corresponds to the Integration Layer of the Boost 4.0 RA that is responsible for connecting the physical entities of the external data sources to the other layers of the Boost 4.0 RA.

Asset Layer

Represents reality, e. g. physical components such as linear axes, metal parts, documents, circuit diagrams, ideas and archives. Human beings are also part of the Asset Layer, and are connected to the virtual world via the Integration Layer. Passive connection of the assets to the Integration Layer, for instance by means of QR codes.

This layer corresponds to the external data sources element of the Boost4.0 RA, which represents the physical entities in the manufacturing environment.

5.3.2 Life Cycle & Value Stream

A further important criterion is the product life cycle with the value streams it contains. This is displayed along the left-hand horizontal axis. Dependencies - such as constant data acquisition throughout the life cycle - can therefore also be represented well in the reference architecture model.

Life Cycle

Industrie 4.0 offers great potential for improvement throughout the life cycle of products, machines, factories and so on. To visualise and standardise Relationships and links, the second axis of the reference architecture model represents the life cycle and the associated value streams. The draft of IEC 62890 is a good guideline for consideration of the life cycle. The fundamental distinction between type and instance is of central importance in those considerations.

a. Type

A type is always created with the initial idea, as a product comes into being in the development phase.

This covers the placing of design orders, development and testing up to the first sample and prototype production. The type of the product, machine, and so on is thus created in this phase. On conclusion of all tests and validation, the type is released for series production.

b. Instance

Products are manufactured industrially on the basis of the general type. Each manufactured product then represents an instance of that type, and, for example, has a unique serial number. The instances are sold and delivered to customers. For the customer, the products are initially once again only types. They become instances when they are installed in a particular system. The change from type to instance may be repeated several times.

Improvements reported back to the manufacturer of a product from the sales phase can lead to an amendment of the type documents. The newly created type can then be used to manufacture new instances. Likewise, each type is also subject to use and updating.

Value Stream

Digitisation and linking of the value streams in Industrie 4.0 provide huge potential for improvement. Links spanning various functions are of decisive importance in this connection. Logistics data can be used in assembly and intralogistics organise themselves based on the order backlog. Purchasing sees inventories in real time and knows where parts from suppliers are at any given moment. The customer sees the completion status of the product ordered during production, and so on. The linking of purchasing, order planning, assembly, logistics, maintenance, the customer and suppliers, and so on, provides great improvement potential. The life cycle therefore must be viewed together with the value-adding processes it contains, and not in an isolated fashion with a view to a single factory, but rather in the collective of all the factories and all the parties involved, from engineering through component suppliers to the customer.

5.3.2.1 Life Cycle & Value Stream in the Boost 4.0 RA

The product lifecycle is displayed in the Business Layer of the Boost 4.0 RA where the various big data enabled applications are deployed to realise business processes. The product lifecycle is also further detailed on the Manufacturing entities dimension where relevant information entities are identified for a phase of the life cycle. These lifecycles cover both the design, production and use of the product. Furthermore, the levels below the business layer incorporate data coming and expose data from different domains such as status of the product ordered during production, order planning, and so on

5.3.3 Hierarchy Levels

The third important criterion, represented in the third (right-hand horizontal) axis, is the location of functionalities and responsibilities within the factories/plants. This represents a functional hierarchy, and not the equipment classes or hierarchical levels of the classical automation pyramid. The third axis of the reference architecture model describes the functional classification of various circumstances within Industrie 4.0. The issue here is not implementation, but solely functional assignment. For classification within a factory, this axis of the reference architecture follows the IEC 62264 and IEC 61512 standards (see figure 2). For a uniform consideration covering as many sectors as possible from process industry to factory automation, the terms “Enterprise”, “Work Unit”, “Station” and “Control Device” were selected from the options listed there and used.

For Industrie 4.0, not only the control device (e. g. head controller) is decisive, but also considerations within a machine or system. Consequently, the “Field Device” has been added below the Control Device. This represents the functional level of an intelligent field device, e. g. a smart sensor. Furthermore, not only the plant and machinery for the manufacture of products is important in Industrie 4.0, but also the product to be manufactured itself. It has therefore been added as “Product” as the bottom level. As a result, the reference architecture model permits homogeneous consideration of the product to be manufactured and the production facility, with their interdependencies.

An addition has also been made at the upper end of the hierarchy levels. The two IEC standards mentioned only represent the levels within a factory. Industrie 4.0, however, goes a step further and also describes the group of factories, and the collaboration with external engineering firms, component suppliers and customers, and so on. For observations beyond the Enterprise level, the “Connected World” has therefore been added.

5.4 IIRA

The IIRA [15] viewpoints are defined by analysing the various IIoT use cases developed by the IIC and elsewhere, identifying the relevant stakeholders of IIoT systems and determining the proper framing of concerns. These four viewpoints are:

1. business
2. usage
3. functional
4. implementation

This document will focus on the functional viewpoint which concerns the functional components in an IIoT system, their structure and interrelation, the interfaces and interactions between them, and the relation and interactions of the system with external elements in the environment, to support the usages and activities of the overall system. These concerns are of particular interest to system and component architects, developers and integrators.

A functional domain is a (mostly) distinct functionality in the overall IIoT system. A decomposition of a typical IIoT system into functional domains highlights the important building blocks that have wide applicability in many industrial verticals. It is a starting point for conceptualising a concrete functional architecture. Specific system requirements will strongly influence how the functional domains are decomposed, what additional functions may be added or left out and what functions may be combined and further decomposed. A typical IIoT system is typically decomposed into five functional domains:

1 Control domain
2 Operations domain
3 Information domain
4 Application domain
5 Business domain

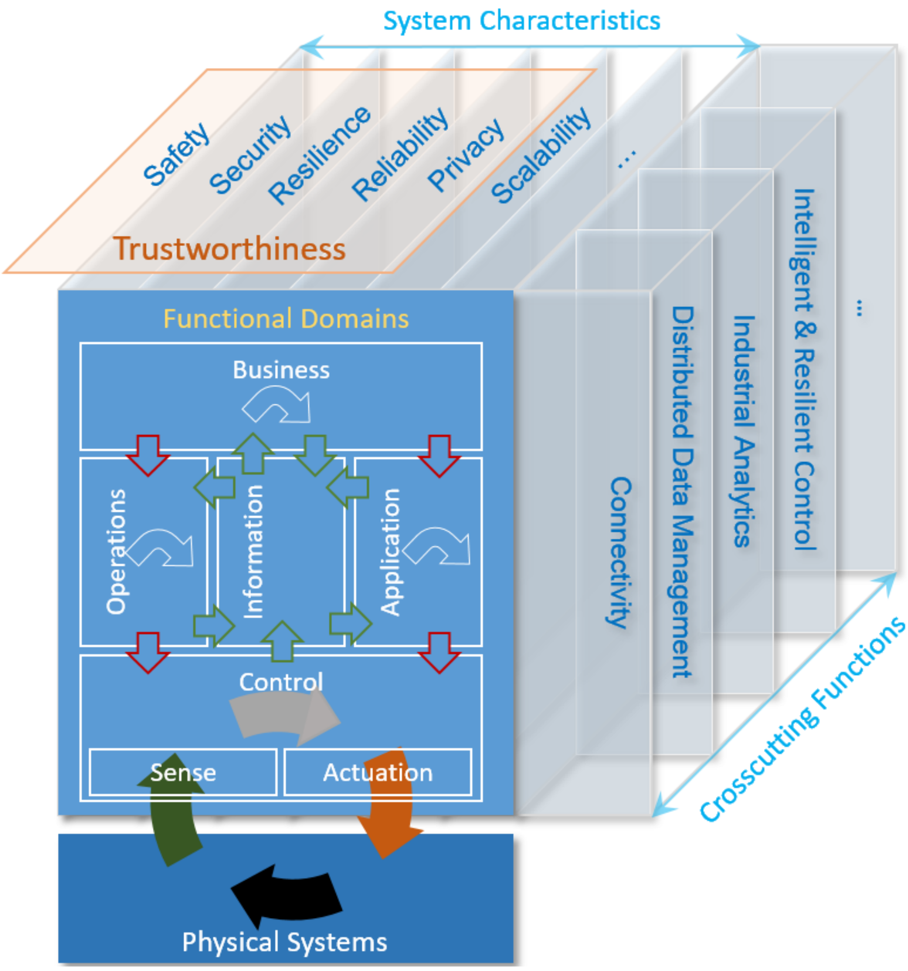

Figure 10 IIRA functional viewpoint

Control Domain

The control domain represents the collection of functions that are performed by industrial control systems. The core of these functions comprises fine-grained closed-loops, reading data from sensors ("sense" in the figure), applying rules and logic, and exercising control over the physical system through actuators ("actuation"). Both accuracy and resolution in timing is usually critical. Components or systems implementing these functions (functional components) are usually deployed in proximity to the physical systems they control, and may therefore be geographically distributed. They may not be easily accessible physically by maintenance personnel, and physical security of these systems may require special consideration.

This Integration layer of the Boost4.0 RA is responsible for connecting the physical entities of the external data sources to the other layers of the Boost 4.0 RA. These data sources may represent among other things sensors, actuators, and so on. Furthermore, control over those entities would also go through the integration layer. Therefore, this layer is roughly equivalent to the Control Domain of IIRA.

Operations Domain

The operations domain represents the collection of functions responsible for the provisioning, management, monitoring and optimisation of the systems in the control domain. Existing industrial control systems mostly focus on optimising the assets in a single physical plant. The control systems of the Industrial Internet must move up a level, and optimise operations across asset types, fleets and customers. This creates opportunities for added business and customer value as set out by higher-level, business-oriented domains.

The functionality incorporated in the Operations Domain (the provisioning, management, monitoring and optimisation of its systems) correspond to either the Application Layer (optimisation) of the Boost 4.0 RA or the Development – Engineering and DevOps cross-cutting aspect (provisioning/management/monitoring).

Information Domain

The information domain represents the collection of functions for gathering data from various domains, more significantly from the control domain, and transforming, persisting, and modelling or analysing those data to acquire high-level intelligence about the overall system. The data collection and analysis functions in this domain are complementary to those implemented in the control domain. In the control domain, these functions participate directly in the immediate control of the physical systems whereas in the information domain they are for aiding decision-making, optimisation of system-wide operations and improving the system models over the long term. Components implementing these functions may or may not be co-located with their counterparts in the control domain. They may be deployed in building closets, in factory control rooms, in corporate data centres, or in the cloud as a service.

Information representation, processing, transformation, and storage is handled in the Information and Core Big Data Layers of the Boost4.0 RA which directly corresponds to the IIRA Information Domain.

Application Domain

The application domain is the collection of functions implementing application logic that realises specific business functionalities. Functions in this domain apply application logic, rules and models at a coarse-grained, high level for optimisation in a global scope. They do not maintain low-level continuing operations, as these are delegated to functions in the control domain that must maintain local rules and models in the event of connectivity loss. Requests to the control domain from the application domain are advisory so as not to violate safety, security, or other operational constraints.

The Application layer of the Boost 4.0 RA groups together components providing pilot case specific logic, and logic supporting the business processes. Additionally, it includes components that facilitate user interaction (input) and services. Its purpose therefore fits in with the Application Domain of IIRA.

Business Domain

The business domain functions enable end-to-end operations of the industrial internet of things systems by integrating them with traditional or new types of industrial internet systems specific business functions including those supporting business processes and procedural activities. Examples of these business functions include Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Product Lifecycle Management (PLM), Manufacturing Execution System (MES), Human Resource Management (HRM), asset management, service lifecycle management, billing and payment, work planning and scheduling systems.

The Business Layer of the Boost4.0 RA combines the functionalities of the layers below to realise business processes and links different business processes in support of the business models of each pilot. Therefore, the Boost.4.0 RA Business Layer corresponds to the RAMI4.0 Business Layer.

Crosscutting functions and system characteristics

The functional components described so far in the functional domains focus on major system functions that are generally required to support generic IIoT usages and to realise generic IIoT system capabilities for business purpose. However, additional functions must be provided to enable the major system functions. These enabling functions, the so-called crosscutting functions, need to be made available across many of the system functional components. For example, system functions need to be connected so they can interact with each other to complete functionality at the system level. Therefore, connectivity is considered a crosscutting function. An important element in the industrial internet is the application of analytics on the data gathered from the industrial assets and control systems to gain insights on their operations. To enable analytics on these asset data, many of the system functional components require concerted effort on data management. Therefore, data management is also considered a cross cutting function. The aggregate behaviour of an IIoT system is not the simple sum of what is provided by its constituent functional components. Like any complex system, there are emergent behaviours or properties resulting from the interactions of the constituent parts. These emergent system-wide properties are called system characteristics.

The system and crosscutting functional analysis largely concern how the system works, while the analysis of system characteristics emphasises how well the system works. For example, to ensure security in a system, a certain set of security functions - as crosscutting functions - must be implemented in each of the functional components and their communications; such as encryption and authentication. However, how secure the system as a whole is depends on how these functions are implemented and how securely these functional components are integrated and interact with each other—as an emergent property: a system is only as secure as its weakest component, or link between components. The same is true for safety or resilience; and any other system property. The trust into a system based on a set of such system characteristics, including safety, security, resilience, reliability and privacy, is defined as trustworthiness of the system.

The realisation of a system characteristic to a certain desired level may depend on, constrain—and in some cases run against—those of other system characteristics. For example, it is not possible to ascertain if a system is safe without also ascertaining if it is secure. An inadequately implemented security measure may be hazardous to safety.

This reference architecture places a strong emphasis on both the functions needed to support the system's business purpose, and ensuring adequate system characteristics so that the functions are performed correctly and the business purpose is not compromised.

Distributed data management and Industrial Analytics cross-cutting functions are directly mapped to the Information and Core Big Data Layers of the Boost 4.0 RA, this is because Boost 4.0 RA focuses on big data. Other cross-cutting functions such as connectivity and/or system characteristics are address by the Boost 4.0 cross-cutting aspects (Data Sharing Platforms, Development – Engineering and DevOps, Communication, Standards, Cybersecurity and Trust.

5.5 Mapping to the Digital Shopfloor Alliance Data-driven Smart Solution Development Framework

The work in WP3 is organised around the Digital Shopfloor Alliance Data-driven Smart Solution Development Framework. As illustrated in Figure 11 it is organised in different levels (Enterprise, Factory, and so on) with each layer consisting of different tools or services and for all of them different modelling approaches are available. A communication pillar (Fog/cloud) enables direct communication between the different levels.

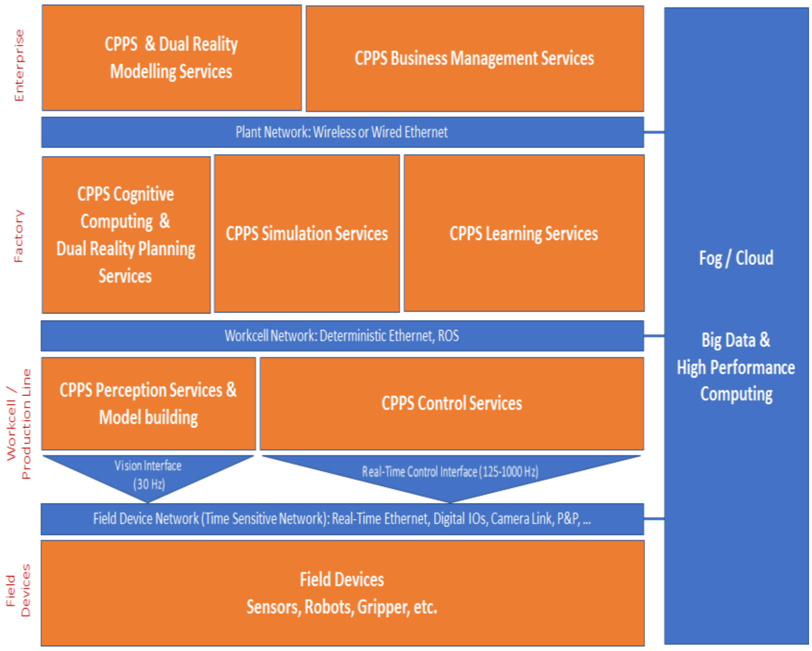

Figure 11 Digital Shopfloor Alliance Data-driven Smart Solution Development Framework

The field Devices level corresponds to the External Data Sources element of the Boost 4.0 RA. The tools and services of the DSA correspond to the Application Layer of the Boost 4.0 RA. However, in order to realise the functionalities in this layer the layers below the application layer are also required. The cross-cutting aspect of communication of the Boost 4.0 RA covers the mechanisms and technologies in the communication pillar. Additionally, the factory dimension indicates which entities these tools and services target. The following figure illustrates an example for the workcell level.

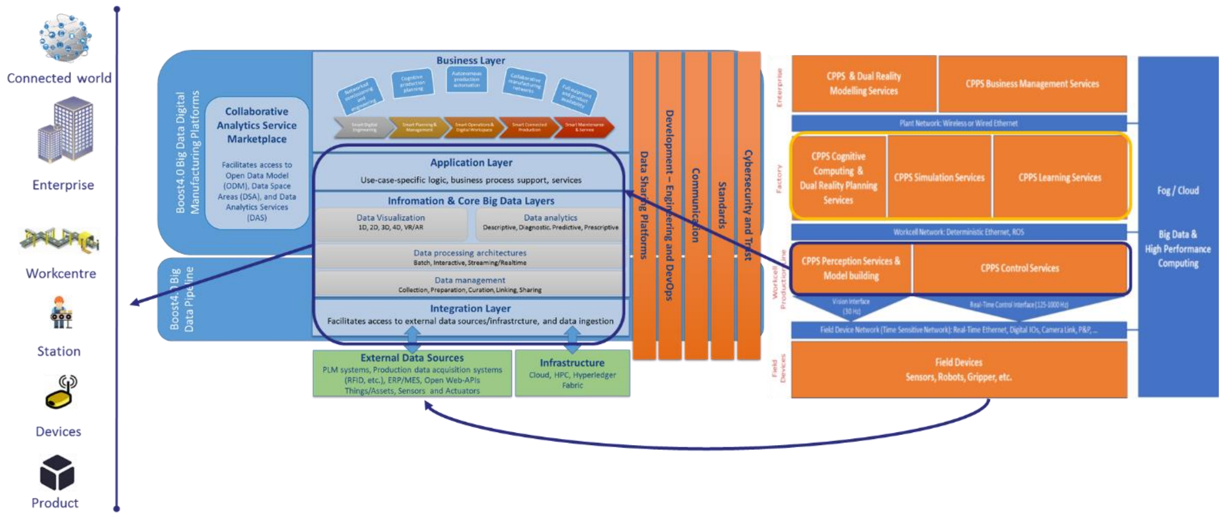

Figure 12 Boost 4.0 RA - Digital Shopfloor Alliance Data-driven Smart Solution Development Framework mapping

5.6 EFFRA Autonomous Factories pathway

The pathway for the smart autonomous factory is shown in Figure 13. The pathway defines an evolution of the digital architectures as well as a digital manufacturing platform, manufacturing equipment and Shop floor automation capabilities.

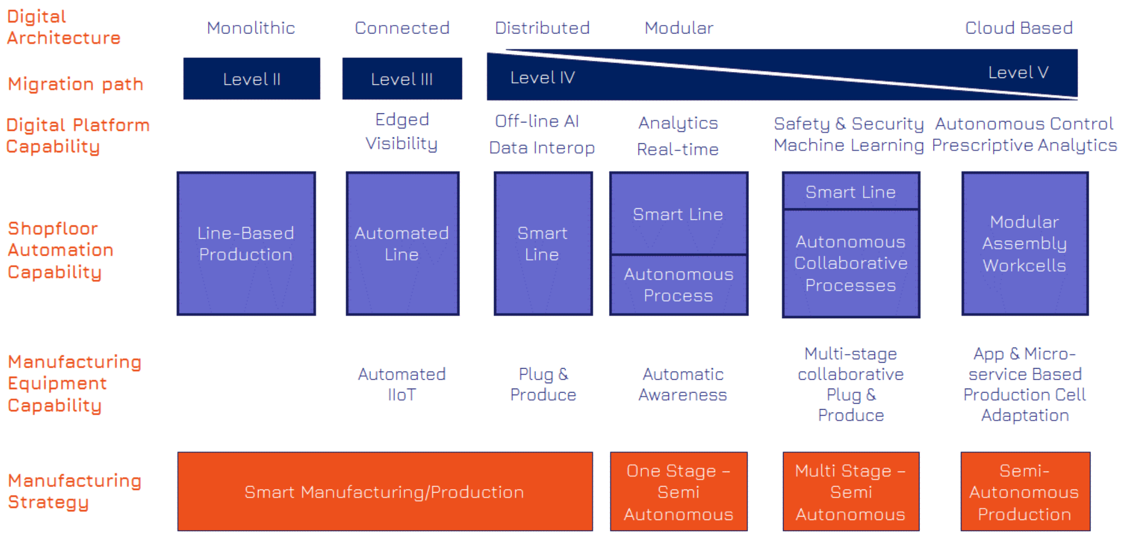

Figure 13: ConnectedFactories smart autonomous factory migration pathway.

Boost 4.0 RA supports Level IV and V of this roadmap in terms of Digital Architecture, covering distributed, modular and cloud-based approaches. Furthermore, Boost 4.0 RA supports offline & online data analytics, and control offered through the application layer, thus realising the digital platform capabilities for levels III-V of the smart autonomous factory pathway.

6 Mapping of IDS, BDE, FIWARE, Hyperledger Fabric

6.1 Industrial Data Spaces

6.1.1 Architecture alignment between IDS and Boost 4.0 RA

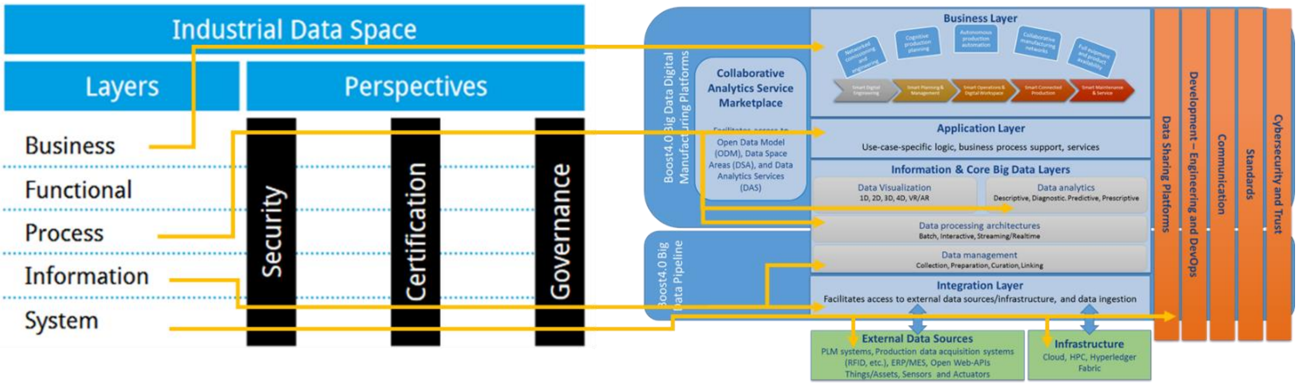

Figure 14 IDS General Structure of Reference Architecture and Alignment to the BOOST 4.0 RA

Data sovereignty is at the core of the IDS, which specifies techniques, methods and processes to enable trustworthy interaction and data exchange patterns based on the IDS Trusted Connectors. These connectors serve as gateways to sensitive data and provide a controlled environment for any kind of data analytics or processing in the form of isolated containers. Therefore, data owners can restrict the usage of their content but at the same time easily cooperate with other participants of the IDS.

The Industrial Data Space initiative proposes a General Structure of the Reference Architecture Model (Figure 14) for this particular capability and related aspects, including requirements for secure and trusted data exchange in business ecosystems. It consists of five layers and three cross-sectional perspectives. The Business Layer specifies and categorises the distinct roles and the main activities and interactions connected with each of these roles. The Functional Layer defines the requirements of the Industrial Data Space. The Process Layer specifies the interactions taking place between the different components and provides a dynamic view of the Reference Architecture. The Information Layer defines a conceptual model using Linked Data principles for describing both data and connectors participating in the IDS. The System Layer considers aspects such as integration, configuration, and deployment of these components.

The perspectives Security, Certification, and Governance affect all layers. Security in the context of the IDS contains all aspects related to a protected and trusted data exchange and usage of Data Apps. The Certification Perspective specifies processes to determine the compliance of participants, both organisations and individuals but also software and hardware components, with the IDS requirements. Governance defines the requirements to be met by the business ecosystem to achieve secure and reliable corporate interoperability. In the following, each layer and perspective is described and how they alignment with the BOOST 4.0 Reference Architecture.

While the IDS Business Layer provides an abstract description of the roles in the Industrial Data Space, it can be considered a blueprint for the other, more technical layers, similar to the BOOST 4.0 Business Layer. The IDS Business Layer can be used to verify the technical architecture of the Industrial Data Space and specify the rights and duties of each participant.

The Functional Layer defines, irrespective of existing technologies and applications, the six fundamental requirements of the IDS, and the features to be implemented resulting thereof. Trusted interactions have several relations with the IDS Roles. Security and Data Sovereignty contain authentication, data access and usage control but also secure and protected communication channels. The Ecosystem of Data regards the challenges associated with discovering and interpreting the provided data. Therefore, every data source in the Industrial Data Space is described based on the Industrial Data Space vocabulary. The technical interoperability is accomplished through the IDS Connectors with their standardised interfaces and exchange protocols. Applications and Data Markets complete the set of functional requirements.

A corresponding element to the IDS Functional Layer is not included in the BOOST 4.0 Reference Architecture. Several parts, regarding security and data management, are covered by corresponding layers and vertical concerns but not in a comparable manner.

The Process Layer specifies the interactions taking place between the different components of the Industrial Data Space, for example data provision and exchange or the publishing of Data Apps. This layer addresses interactions and business processes as regarded in the BOOST 4.0 Application Layer but also data processing

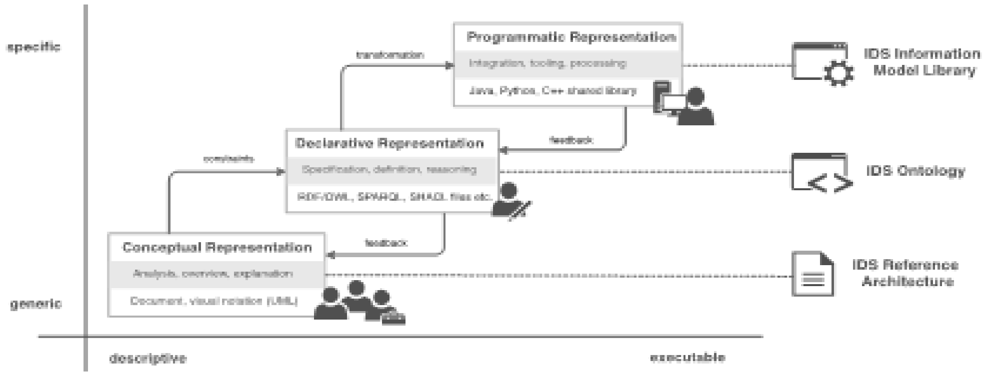

Figure 15 Representations of the IDS Information Model

and management tasks as described by the BOOST 4.0 Data Layer.

The IDS Information Layer specifies the Information Model, defining the domain-agnostic lingua franca of the Industrial Data Space. The Information Model constitutes a central agreement shared by its participants and components, facilitating compatibility and interoperability; with no commitment to any particular domain. It has been specified at three levels of formalisation where each level corresponds to a digital representation, ranging from a high-level, conceptual document up to the level of operational code (Figure 15). At its core is the IDS Ontology [111] as a normative specification of the Information Model in RDF. Its main purpose are the uniform description of data and software assets and their integration therefore covering the Data Management and Integration Layer of the BOOST 4.0 Reference Architecture.

The System Layer roles specified on the Business Layer are mapped onto a concrete data and service architecture in order to meet the requirements specified on the Functional Layer, resulting in what is the technical core of the Industrial Data Space. Its specifications regarding integration of both internal and external data sources as well as providing secure communication channels align with various elements of the BOOST 4.0 Reference Architecture. The basic element described in the System Layer is the IDS Connector as a common, standardised gateway to the Industrial Data Space itself. The IDS Trusted Connector (Figure 16) serves as a specialised version with full security and data protection capabilities. As such, the IDS Trusted Connector also meets the requirements of the BOOST 4.0 sections Communication, Data Sharing Platforms, and Cybersecurity and Trust out of the box.

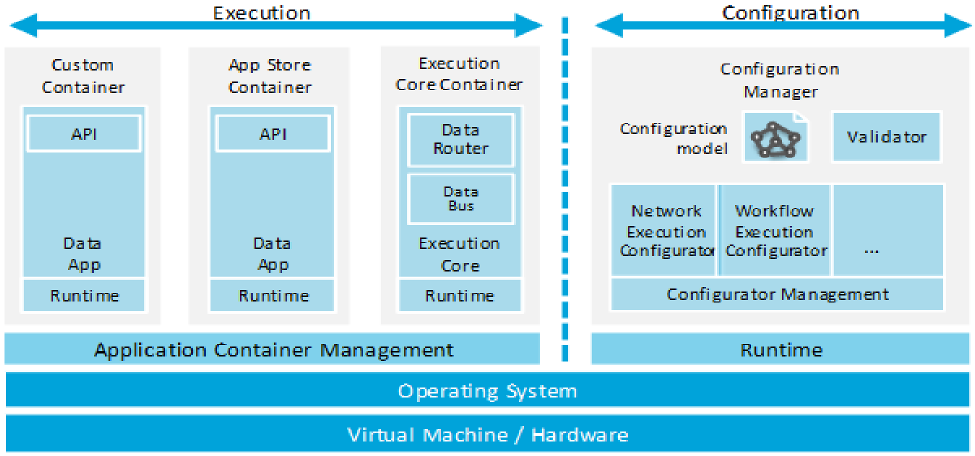

Figure 16 Components of the IDS Trusted Connector

The Security Perspective provides means to identify participants, protect communication and data exchange transactions between them, and control the use of data after it has been sent, in particular in combination with the Trusted Connector and across all layers of the architecture. To the extent possible and reasonable, existing standards and best practices are to be used and leveraged, for example ISO/IEC 27000 or ISO/IEC 9594-8.

Together with the IDS Certification processes, the Security specification pave the way to a secure and trustworthy network. The obvious similarities with the BOOST 4.0 Cybersecurity and Trust view allow a direct integration of the IDS Trusted Connector into the respective section of the BOOST 4.0 Reference Architecture.

The IDS Reference Architecture Model consequently covers the major parts of the BOOST 4.0 elements but with a strong focus on security, data protection and sovereignty. The broader view of the BOOST 4.0 Architecture enhances the IDS architecture especially in the context of Big Data applications and its specifications regarding data analytics. Together, they form a comprehensive set of concepts and views to describe applications and networks of Industrial IoT systems.

6.1.2 Mapping of IDS on the Boost 4.0 RA

6.1.2.1 IDS Framework

The most straightforward use of IDS is as a Data Sharing Platform that facilitates data exchange between the various layers of the Boost 4.0 RA. As such the IDS framework is mapped directly Data Sharing Platform cross-cutting aspect of Boost 4.0 RA. However, IDS framework also provides functionalities for data sovereignty and protection and therefore it also supports the Cybersecurity and Trust cross-cutting aspect.

6.1.2.2 IDS Appstore

The IDS Appstore may be used to share IDS applications and connectors.

6.1.2.3 IDS applications

The IDS framework enables the development of applications integrated within the IDS connectors and provides a variety of functionalities these applications based on their functionality could be placed on different layers.

6.1.2.4 IDS example mapping

The following figure illustrates how IDS framework components maybe mapped on the Boost 4.0 RA.

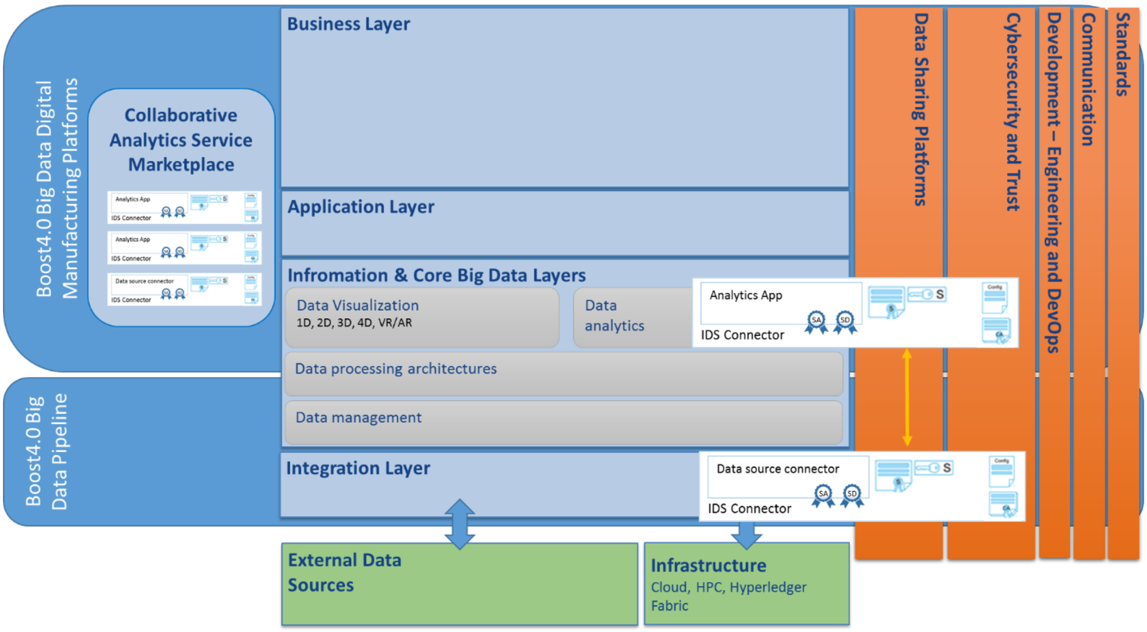

Figure 17 IDS mapping on Boost 4.0 RA

6.2 Big Data Europe

6.2.1 Architecture alignment between BDE and Boost 4.0 RA

The Big Data Europe is an easy-to-deploy, easy-to-use and adaptable (cluster-based and standalone) platform that leverages easy installation and execution of big data tools like Apache Hadoop, Apache Spark, Apache Flink, Apache Flume and Cassandra.

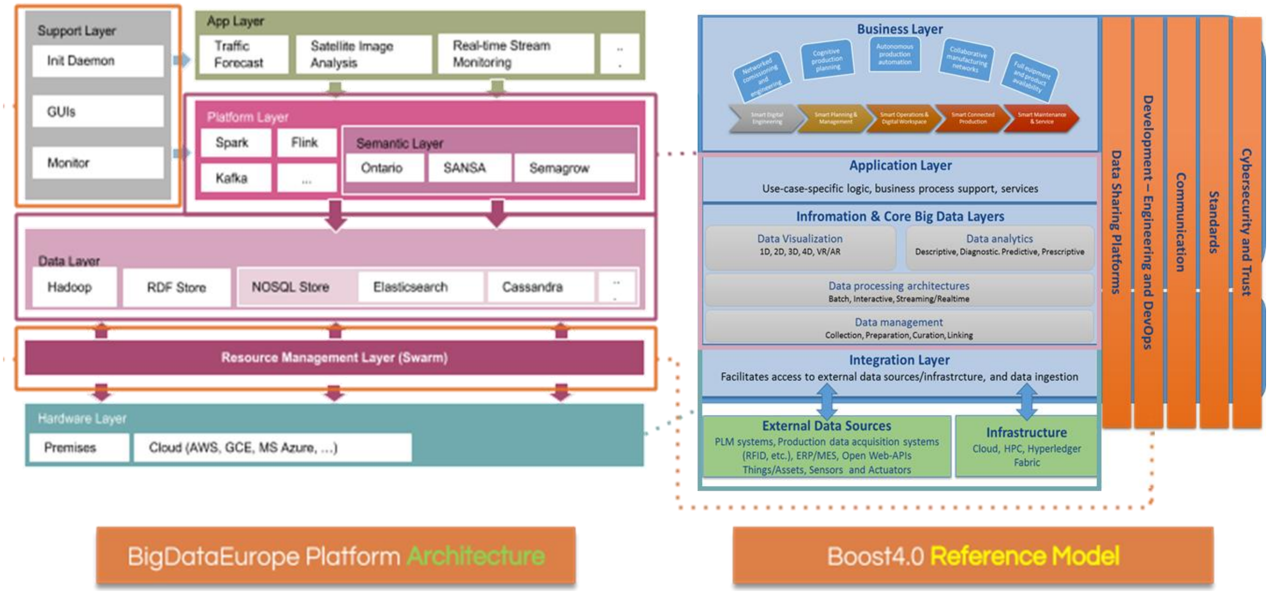

Figure 18 BDE Platform architecture overview and its alignment with the Boost 4.0 RA

Figure 18 provides a high-level overview on the BDE platform architecture, which is elaborated below:

Hardware layer and Resource management Layer

BDE uses Docker [95], as a lightweight packaging and deployment methodology in order to easily manage the variety of the underlying hardware resources efficiently. BDE provides a large number of dockerised Big Data components [96], mainly by providing base Docker images [97]. BDE uses Docker Swarm as a clustering and scheduling tool for Docker containers by establishing and managing a cluster of Docker nodes as a single virtual system. Docker Swarm acts as the resource manager and it can help orchestrate the Docker components on the cluster of heterogeneous machines. To complement both, Docker compose helps to create a stack of multiple simultaneous running Docker containers. These layers could be mapped directly into the Boost 4.0 reference architecture, in particular at the Infrastructure Layer and Integration Layer.

Support layer

The technical aim of the BDE Platform is to reuse components wherever possible and build tools that are necessary to fit the user's need. As many applications and services depend on each other, the start-up may fail when all the services will be starting at once. This is not always the intended behaviour, since some applications may depend on each other, or on a human intervention. It uses a so-called init daemon service [98] that, given an application-specific workflow, orchestrates the initialisation process of the components.

This service may be mapped at the Boost 4.0 architecture at the level of Development - Engineering and DevOps.

User interfaces

To lower the usage barrier of Big Data technologies, BDE has developed a set of UIs, including a stack builder [99], workflow builder [100], a pipeline monitor [101], an integrator UI [102], and a Swarm UI [103]. The UIs serve different purposes, but in general make it easier for the user to build, deploy and monitor applications on the BDE platform. All these user interfaces could be aligned and used on the Boost 4.0 Application Layer which then could be adapted by the Business Layer and application specific needs.

Semantic Layer

Big Data involves three classic dimensions volume, velocity and variety. While the former two are well supported by a plethora of software components, the variety dimension is still rather neglected. One of the prominent methods to deal with the heterogeneous data sources are knowledge graphs. As an add-on to the BDE platform, an open-source Semantic Analytic Stack (SANSA) framework [17] has been introduced and developed. It allows scalable analysis of large-scale knowledge graphs to facilitate applications such as link prediction, knowledge graphs completion and reasoning. The motivation behind this work lies in the lack of scalable methods for analytics which exploit expressive structures underlying semantically structured knowledge bases. It is based on the BDE technical platform, which utilises Docker technology. The Semantic Layer from the BDE platform, in particular SANSA framework, could be beneficial for covering heterogeneity between stakeholders and data providers for better and efficient Data Processing, Data Management and Data Analytics.

6.2.2 BDE component mapping to the Boost 4.0 RA

BDE dockerised components based on their functionality are mapped as follows:

- The Data management layer covers the following BDE components:
    - Apache Hadoop
    - Virtuoso
    - 4Store
    - Apache Cassandra
    - Apache Hive
    - Silk
    - SemaGrow
    - Strabon

- Ontario
- SANSA

➢ The Data processing architectures covers the following BDE components:
  - Apache Spark
  - Apache Hadoop
  - Apache Flink
  - UnitedViews
  - Apache Flume
  - Apache Kafka

➢ The Data analytics layer covers the following BDE components:
  - FOX
  - GeoTriples

➢ The Data Visualisation layer covers the following BDE components
  - Sextant
  - Hue

➢ The Development – Engineering and DevOps aspect covers the following BDE components:
  - Big Data Integrator

Figure 19 illustrates the mapping:

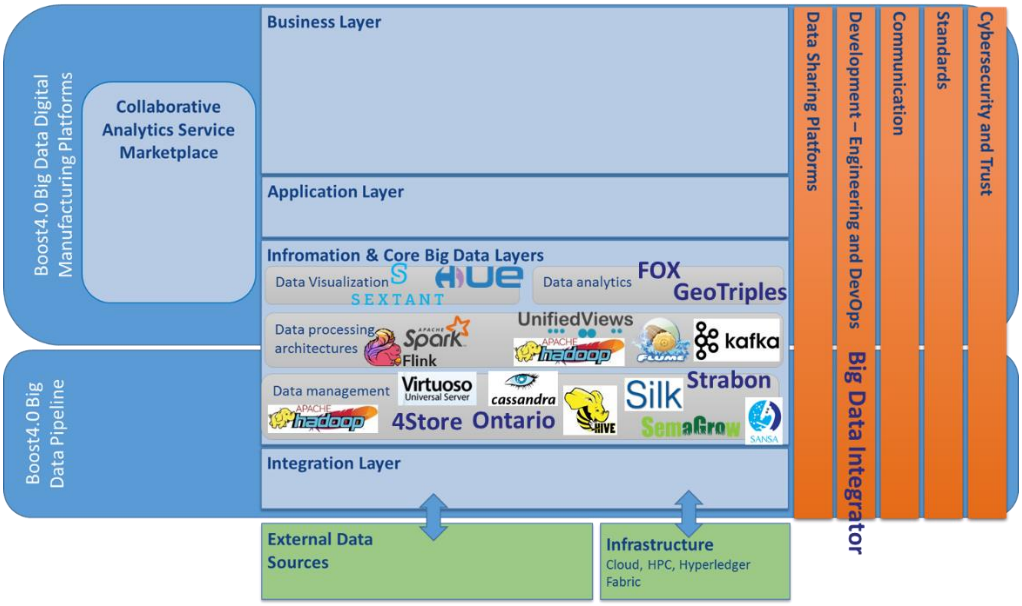

Figure 19 BDE components mapping

6.3Hyperledger Fabric

At a technical level, a blockchain can be defined as an immutable ledger for recording transactions, kept within a distributed network of mutually untrusting peers. Every peer keeps a copy of the ledger. The peers execute a consensus protocol to validate transactions, group them into blocks, and build a hash chain over the blocks. This process forms the ledger by ordering the transactions as is necessary for consistency [19].

Blockchain frameworks typically include the following four building blocks:

A shared ledger: Append-only distributed system of record shared across business network. The ledger is immutable and final. Immutability means no participant can tamper with a transaction once it is agreed. Finality means that there is one place to determine the ownership of an asset or completion of a transaction. This is the role of the shared ledger. *Assets* are anything of value. An asset can be tangible or intangible. It has a state and ownership.

Cryptography: Cryptography in a blockchain ensures authentication and verifiable transactions. Cryptography is used to ensure that network participants see only the parts of the ledger that are relevant to them, and that transactions are secure, authenticated, and verifiable.

Consensus: Mechanism through which network participants agree how transactions are verified. All parties agree to network verified transaction before the transaction is written to the ledger.

Business rules or smart contracts: Smart contracts are the business terms that are embedded in a blockchain transaction database and executed with transactions. This is also the rules component of a blockchain solution. It is needed to define the flow of value and state of each transaction.

How blockchain works

- Every party has same view of asset state
- Client constructs transactions and broadcast to network
- Network nodes process the transaction (smart contract)
- Network nodes agree on the output/result (consensus)
- Network nodes save the transaction on the immutable ledger
- Repeat

Hyperledger Fabric (Fabric), is a permissioned blockchain infrastructure (for a comparison between permissioned and permissionless models refer to [20]), providing a modular architecture with a delineation of roles between the nodes in the infrastructure, execution of smart contracts (called "chaincode" in Fabric) and configurable consensus and membership services.

The Linux Foundation Hyperledger Fabric [104] has value-added enterprise ready functionality such as:

- Permissioned membership
- Performance, scalability, and levels of trust
- Data on a need-to-know basis
- Rich queries over an immutable distributed ledger
- Modular architecture supporting plug-in components such as security and identity
- Protection of digital keys and sensitive data

Figure 20 shows a high-level overview of applications in Fabric. The blockchain developer writes the client application and the smart contract (chaincode). The latter can be written in Go; Node.js (v1.1); and Java (v1.2). Client SDK enables the creation of applications that deploy and invoke transactions atop a shared ledger. The Hyperledger Fabric Reference Architecture supports both Node.js and Java SDK. The application SDK transmits the transaction to the ordering service. Ordering service is a communication service run by a group of *Orderers* to provide atomic broadcast. Orderer manages a pluggable trust engine that performs the ordering of the transactions. Clients of ordering service are peers and applications. A *Peer* is a node on the network maintaining state of the ledger and managing chaincodes. Any number of Peers may participate in a network. A Peer can be an endorser, committer. An endorser is always a committer. An endorser executes and endorses transactions. A committer verifies endorsements and validates transaction results. A Peer manages event hub and deliver events to the subscribers. Peers form a peer-to-peer gossip network. When information related to data on chain is requested, the client application queries the ledger to extract the required information.

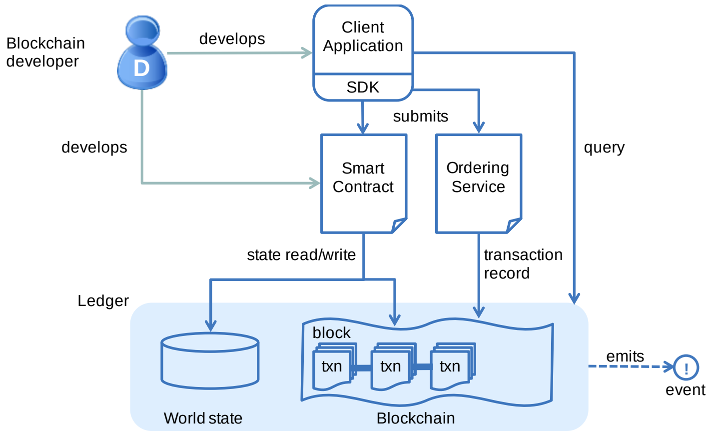

Figure 20: Hyperledger Fabric Applications Overview

Generally speaking, a Fabric Network comprises *Peer nodes*, which execute chaincode, access ledger data, endorse transactions, and interface with applications, *orderer nodes* which ensure the consistency of the blockchain and deliver the endorsed transactions to the peers of the network, and *MSP services*, generally implemented as a Certificate Authority, managing X.509 certificates which are used to authenticate member identity and roles. The Fabric network is mapped to the data management layer in BOOST 4.0 RA. The ledger itself is the underlying infrastructure of blockchain applications and therefore mapped to the infrastructure layer. The client application comprises the business logic of the application and therefore mapped to the business layer, while the SDK as development tool is mapped to the development vertical layer. Figure 21 shows how the different components in Fabric are mapped to BOOST 4.0 RA.

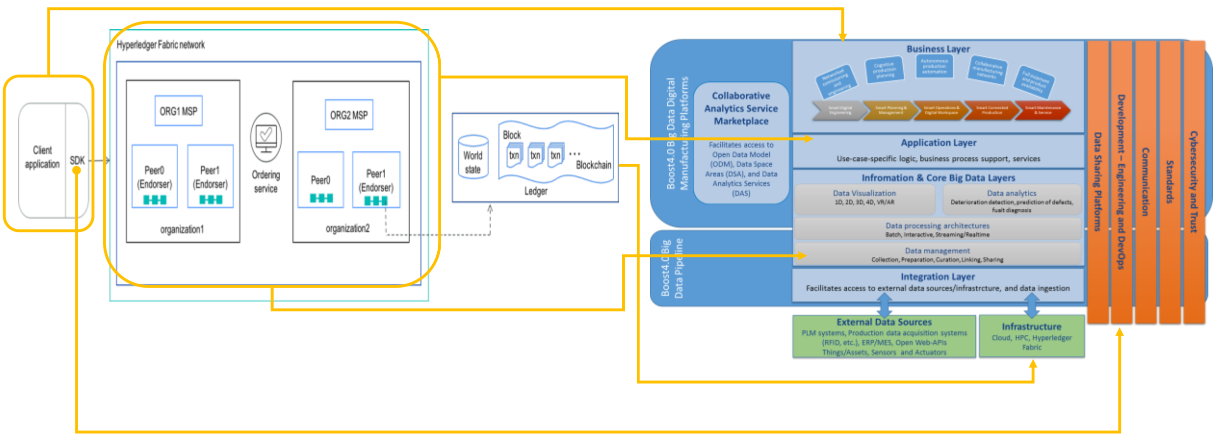

Figure 21 Mapping of Hyperledger Fabric to BOOST 4.0 RA

7 Boost 4.0 RA instantiations

This section makes a connection between the Boost4.0 trials and the Boost4.0 RA. It is organised per trial and provides a short overview each trial in terms of business process and the components that make-up the existing solutions. The components are further detailed and associated with each trial's system requirements in D2.1 – Pilot Requirements and Use Cases Specification v1. Additionally, this paragraph documents how the solutions of each trial fit into the Boost4.0 RA and discusses component integration.

7.1 Volkswagen Light Casting Metal

7.1.1 Short pilot overview

The Volkswagen pilot will implement a set of interconnected processes that will boost the effectiveness of production, reduce costs and increase the quality of produced light metal casting parts. Following business processes have been elaborated to achieve these goals:

- *Real-time continuous tool condition monitoring.* Automated experience and simulation-based approach using sensor data and big data analysis methods potentially transferable to other production processes.
- *Simulated tool-health prediction models:* The mould health state needs to be predicted for process parameter adjustment or immediate tool maintenance
- *Integrated mould design:* Big data from mould usage and data from external plants and suppliers needs to be combined and analysed with production data (for example product geometry, mould temperatures, acoustic sensors, thermal and optical sensors and imaging) and mould sensor data for better mould commissioning.
- *Smart product data.* Engineering modules that will integrate and analyse "smart vehicle big data" as the feedback loop to improve automotive part quality, mould design processes and production cost-efficiency.

7.1.2 Short component description

VW Pilot component	BOOST 4.0 RA	Comment
VW Assistant Systems	Application Layer Business Layer	The Volkswagen Assistant systems shall use any kind of available data and data analysis results to provide useful assistance features to actors

	Development, Engineering and DevOps	involved in the light metal casting process of Volkswagen. The features provided by assistant systems shall address functionalities located in the application and business layer of the BOOST 4.0 RA as well as features for development, engineering and DevOps.
ESI casting simulation suite	Application Layer Business Layer Development, Engineering and DevOps	The ESI casting simulations suite provides features for simulation synchronised implemented light metal casting process devices located in the shop floor. As well as Volkswagen assistance systems, the ESI casting simulation suits shall provide features located in the application and business layer of the BOOST 4.0 RA as well as features for development, engineering and DevOps.
ESI INENDI Inspector	Application Layer Business Layer Development, Engineering and DevOps Data Visualization Data Analytics	The ESI INENDI Inspector offers features for multi correlation analysis to be used in Volkswagen. In relation to the BOOST 4.0 RA: It addresses the application and business layer in case of further extension within the Volkswagen pilot. It addresses the domains for Development, Engineering, DevOps and mainly analytics and visualization of data.
EIDS Connector	Integration Layer Data processing architectures Data Management	The VW pilot uses the EIDS for B2B information exchange. The EIDS connector provides next to the integration of systems into the EIDS also features to process and manage data. Therefore, the EIDS connector addresses the integration layer, Data processing architectures and Data Management of the BOOST 4.0 RA.
OPC UA integrator	Integration Layer Data processing architectures Data Management	The VW pilot uses the proprietary OPC UA service bus ICT infrastructure for information exchange and the OPC UA integrator to connect systems to this service bus. The OPC UA integrator provides next to the integration of systems into the Volkswagen ICT

		infrastructure also features to process and manage data. Therefore, the OPC UA integrator addresses the integration layer, Data processing architectures and Data Management of the BOOST 4.0 RA.
EIDS	Infrastructure Collaborative Analytics Service Marketplace Data Sharing Platforms Communication Standards Cyber Security and Trust	Volkswagen sees in future potential to establish a B2B communication with selected companies. The EIDS should therefore cover many aspects of the BOOST 4.0 RA. It should provide a platform infrastructure, it should provide a collaborative Analytics Service Marketplace for data pre-processing possibilities, and it should enable the sharing of data, through secured and trusted communication valid and trustful through the use of standards.
OPC UA Service Bus	Infrastructure Data Sharing Platforms Communication Standards Cyber Security and Trust	The VW pilot uses the proprietary OPC UA service bus ICT infrastructure for internal systems communication. As the EIDS also the Volkswagen ICT infrastructure already covers a lot of aspects of the BOOST 4.0 RA. The ICT infrastructure runs on a Volkswagen internal server infrastructure. It enables secure and trustful data sharing and communication between connected clients in a standardised way.
Smart Moulds	External Data Sources	The VW pilot uses Smart Moulds as data sources.
FILL Machines	External Data Sources	The VW pilot uses FILL machines as data sources.
Vehicles	External Data Sources	The VW pilot intend to use Vehicle data as data sources.

7.1.3 Mapping to the RA

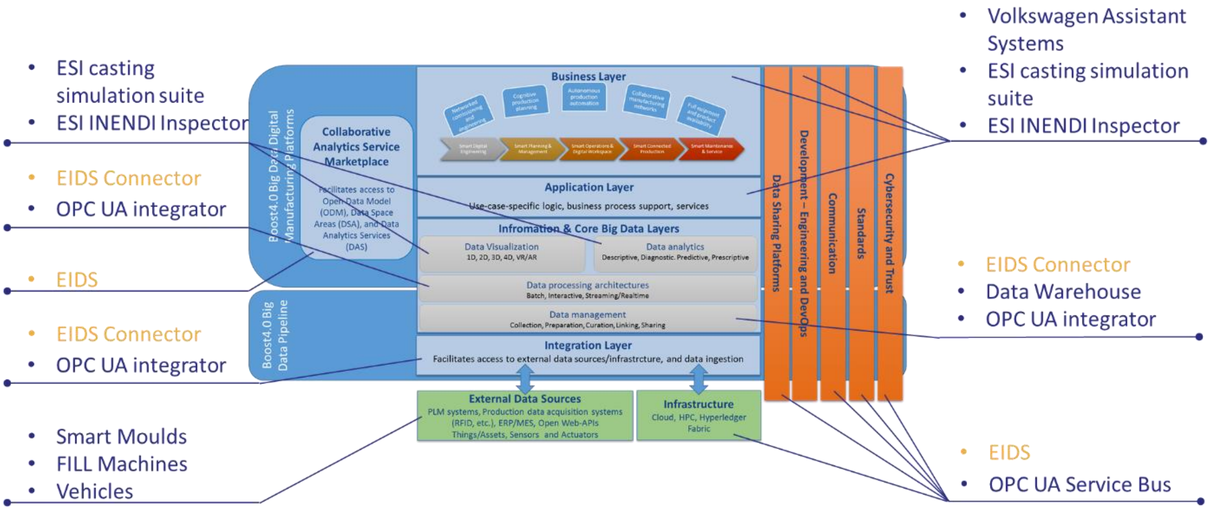

Figure 22 Big Data Pilot Mapping to Boost 4.0 Reference Architecture

Figure 22 shows the mapping of the VW pilot to the BOOST 4.0 Reference Architecture (RA). The VW pilot addresses all parts of the RA as explained in the following.

- External Data Sources – The VW pilot uses Smart Moulds, FILL machines and Vehicle data as data sources.
- Infrastructure – The VW pilot uses the proprietary OPC UA service bus ICT infrastructure as well as the EIDS for B2B information exchange.
- Integration Layer – To facilitate access to data sources and Infrastructures, the VW pilot uses the OPC UA integrators for the proprietary OPC UA services bus infrastructure and the EIDS connector for the EIDS integration.
- Data Management – The VW pilot uses data warehouses connected to the OPC UA service bus, as well as the connectors and integrators to the OPC UA service bus and the EIDS for the data management functionalities as collection, preparation, curation, linking and sharing of data.
- Data processing architectures – The VW pilot uses the EIDS connector and the OPC UA integrator for data processing purposes as batch and streaming data exchange.
- Data Visualisation – The VW pilot uses the casting simulations suite and the INENDI inspector provided by ESI for data visualisation purposes.
- Data Analytics – The VW pilot uses the casting simulations suite and the INENDI inspector provided by ESI for data analytics purposes.
- Application Layer – The VW pilot implements use-case specific logic and services in Volkswagen Assistant Systems also derived from the ESI casting simulation suite and the INENDI inspector.

- **Business Layer** – The VW pilot provides relevant business information in Assistant Systems, the ESI casting simulation suite and the INENDI inspector.
- **Collaborative Analytics Service Marketplace** – The Service Broker as component of the EIDS will cover this aspect.
- **Data Sharing Platforms, Communication, Standards, Cyber Security and Trust** – The EIDS and the OPC UA Service Bus are providing these features or covering aspects in the VW pilot.
- **Development, Engineering and DevOps** – These aspects are addressed by Volkswagen assistant systems, the ESI casting simulation suite and the INENDI inspector in the VW pilot.

7.1.4 Integration of components

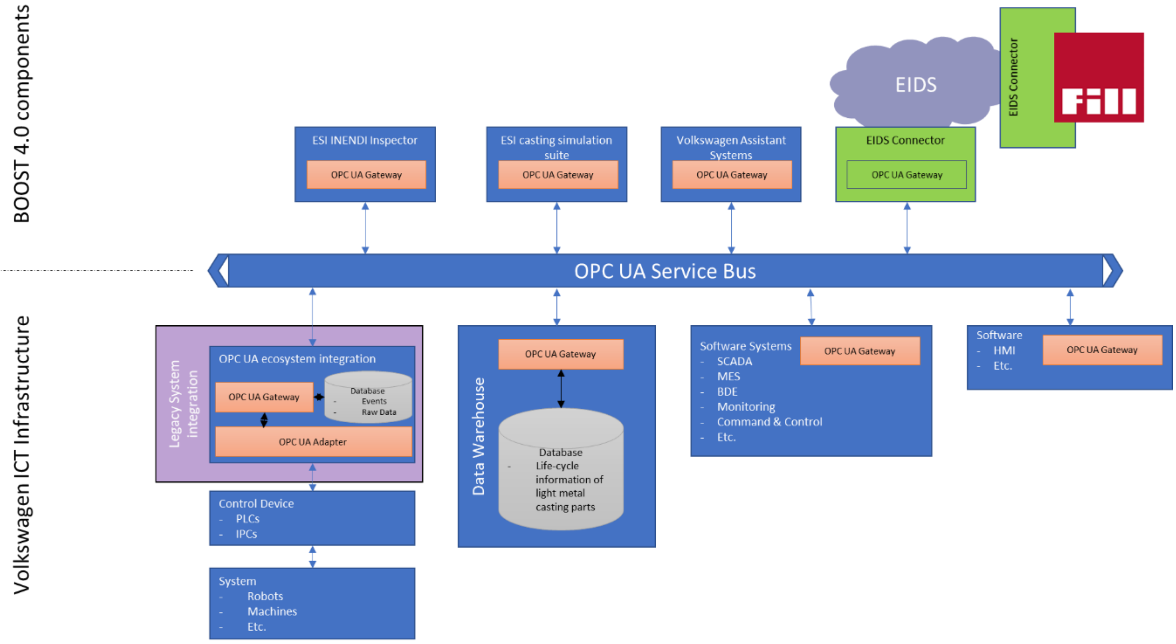

Figure 23 ICT architecture in Volkswagen

Current projects in Volkswagen transform today's brownfield light metal casting processes into an Industrie 4.0 compliant data space. Results of these brownfield transformation projects shall be basis of the VW pilot in BOOST 4.0.

The lower part of Figure 23 shows conceptually the currently running brownfield transformation in the Volkswagen light metal casting process. An OPC-UA service bus enables cross-level connectivity between legacy systems, data warehouses, software systems and further components as HMI, and so on Legacy systems without OPC-UA interface gets integrated by an OPC-UA ecosystem integration module. This module provides firstly an OPC-UA adapter which transforms and unifies proprietary legacy system interfaces into an OPC-UA compliant communication protocol; secondly a database which saves measurements data and events (enables also pre-calculation and aggregation possibilities), and finally an OPC-UA interface which connects the module to the OPC-UA service bus. OPC-UA enabled systems are connected to the OPC-UA service bus via an OPC-UA gateway.

The upper part of Figure 23 shows the integration of components to be developed within the BOOST 4.0 project. ESI provides data analytics tools as the ESI casting simulation suite and the INENDI Inspector. These tools are connected by an OPC-UA gateway to the Volkswagen OPC-UA service bus to enable connectivity with the Volkswagen legacy systems and Volkswagen assistant systems in production. The Volkswagen assistant systems are preparing data analytics results coming from ESI tools into a format which are represent useful information for different actors.

The OPC UA Service Bus could be connected to EIDS gateway to enable B2B information exchange (for example) with FILL to exchange machine data with their vendor to enable remote maintenance and optimisation and to improve the machine design on vendor side (feedback to design).

7.2 FILL Gurten

7.2.1 Short pilot overview

The **FILL pilot** is about a lot-size-one machine tool circular engineering in a factory 4.0. FILL as key machine tool provider delivers in the order of 100 production systems per year to around 50 customers worldwide. 90% of FILL machines and systems are exported to automotive, aerospace, sports, construction and housing and renewable energy sectors. Zero defect production demands that machines cost-effectively and flexibly adapt to optimum production configurations. Therefore, Machine tools are increasingly customised-customisable (lot-size 1 production scheme). Rigid engineering processes designed for mass production are not able to optimise smart connected machine tool lot-size 1 engineering and fail to incorporate external operational data to optimise machine self-configuration and adaptation features.

The FILL pilot follows the following objectives:

1. <u>Long-term reduction of machine development times</u>

First to market with the leading edge technology on smart engineering on industrial big data, FILL will become the leading smart machine builder. Reducing time-to-market of innovative customised products is a key success factor for industrial companies. Integrating big data feedback information from operation and maintenance phases into the engineering phases will shorten the time for real plant or factory commissioning in lot-size-1 production facility. With regard to the life cycle, this consideration also includes optimisation potentials with regard to delayed requirements and rework. All this aims to strengthen the urge to transform towards an agile organisation.

2. <u>Establish pattern and anomaly detection framework</u>

In the new proposed engineering methodology for model-based and big data-driven engineering processes it is essential to establish a pattern and anomaly detection framework (connected to the results of WP2-WP3). With this framework different behaviour models as well as artificial intelligent and machine learning algorithms are developed. The models are used in the engineering process to get better insights of the physical, logistics or other behaviours. This will accelerate the decision-making process, reduces or accelerate iterations in all domains of the engineering and therefore improvements in the sense of time, quality and costs. The models are stored in a model repository and can be further used by the customer, for example as virtual sensors. This could lead in a new business model "Model as a Service". Therefor this objective is also connected to the global Boost 4.0 Objective **O3-Maximise BOOST 4.0 Commercial Impact**.

3. <u>Integration and synchronisation of machine, product and process data</u>

In the early phases of the engineering process, mainly in the conceptual phase (smart digital MBSE), usually many data sources are required in order not to overlook any boundary conditions and requirements. The integration and synchronisation of different data coming from different data sources is an enabler for establishing the smart digital engineering process using big data. In sense of Big Data Analytics this will improve the informative value of the analysis output, the quality of the developed models and increases the model diversity. This needs knowledge from different ICT topics like interfaces (OPC UA, MQTT, RT), data management (sql, NoSQL, hybrid) and so on. This objective is based on the results of WP3.

4. <u>Simulation - PLM integration</u>

To use the large number of CAD-Models efficiently, interfaces between Visual Components simulation tools and the engineering management tools of Fill have to be established, thus a better integration of the simulation process into the proposed engineering process will be achieved. This leads to a simulation management in which saving, loading and version control are well integrated. Moreover, design changes will be updated faster in the simulation. The target from VIS is to create a generic PLM interface and tailored with add on for the PLM solutions in the market.

5. <u>Developing of Digital Twins</u>

Depending on the application and task, digital twins at different level of detail and granularity will be developed. However, the main objective will be to map the different states of the production system or parts of it over the life cycle. In this case the solution will focus on three states

- as-engineered (state after the engineering was finished)
- as-manufactured (state after manufacturing and in-house commissioning)
- as-operated (how the customer operates the production system, simulate historical real data)

In order to extend the field of application from pure simulation and monitoring usage, the requirements have to be integrated into the digital twin. This enables,

- requirements analysis for new projects
- failure analysis by comparing the three states
- decision making during the engineering process

6. <u>Customer centric outcome</u>

The development of innovative, highly customised production systems is generally based on a customer-centric approach. Furthermore, for the machine builder, the operations process knowledge is a key success factor. This is mostly empirical knowledge and is usually built up during the operating phase and used in follow-up projects. The intended feedback loop should accelerate the build-up of knowledge and make it possible to secure customer wishes earlier and increases customer satisfaction. This leads to fewer delays in the business process like the customer dependent approval processes, for example design approval, delivery approval and so on and in the following to earlier cash flows which are linked via a payment schedule to the approval process.

7. <u>Human resource development</u>

A key factor in the implementation of such far-reaching strategic business objectives are the right employees. Especially in the traditionally conservative area of special machine builders, new methods and technologies are only slowly gaining acceptance. Moreover, a machine builder is not seen as a potential employer on the ICT job market, so finding ICT talents is very difficult. Actions must be taken to increase the company's visibility to ICT talents (developing recruiting strategy).

Actions:

- State-of-the-art advertisements for IoT-Architect, Data Scientist and so on.
- Marketing measures at trade fairs (extra area for ICT), in journals and so on.
- Offering Master and Bachelor programmes of studies in these areas of research.

7.2.2 Short component description

The software architecture of Fill's IoT platform (the official product name is Cybernetics) (see Figure 24) follows a layered approach.

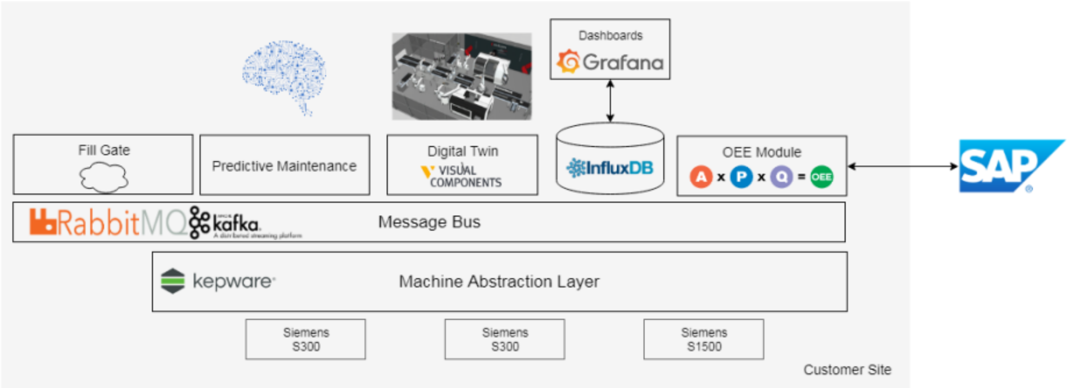

Figure 24 FILL's IoT Platform

The first layer is the hardware layer. The hardware layer contains all physical devices (for example PLCs, sensors …) that produce machine or production data that is eventually collected by Cybernetics.

Fill is a lot-size one machine builder, which results in a very heterogeneous landscape of different machines, sensors, and data sources. Hence, the machine abstraction layer is an essential part in the Cybernetics architecture. The machine abstraction layer collects data from different sources and converts it to a unified format. This data format can be processed by any component built on top of Cybernetics.

On top of the machine abstraction layer sits the data distribution layer/message bus, which is implemented using a message broker. Any data point, collected from any data source is distributed among all components of Cybernetics via the message bus. Broadcasting all data on a message bus allows implementing different data consumers and a plug-and-play-like approach for adding and removing components of cybernetics.

Finally, different modules of the application layer consume data from the message bus. Cybernetics provides a data storage module, which simply stores all data objects on the message bus in various databases (for example time series databases, document storage …). These databases allow other Cybernetics modules accessing history data at any time.

Cybernetics can also integrate with other software platforms (for example ERP systems) and import/export data or drive various analyses (for example the OEE module).

The predictive maintenance module of cybernetics uses machine learning to; for example, analyse the wear of ball screws as part of the milling machines. Finally, the digital twin plug-in of Visual Components allows a live visualisation of the current machine state (for example motion visualisation).

The Fill Gate is a module that allows transferring data of Fill machines (customer site) back the Fills data lake (vendor site). Fill then uses this data to provide software services or enhance engineering.

7.2.3 Mapping to the RA

The architecture of Cybernetics directly maps to the Reference Architecture, as shown in Figure 25.

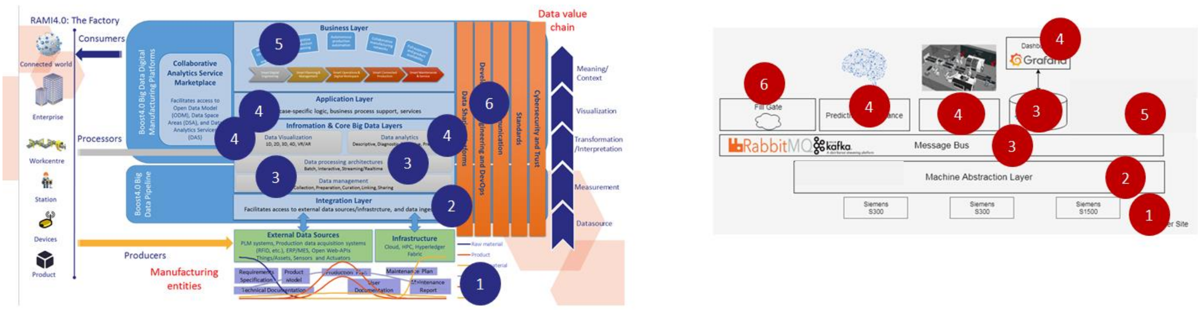

Figure 25 Mapping the Cybernetics' Architecture to the Reference Architecture

On the bottom of the RA there are the *manufacturing entities*. In the case of Fill, these entities are different PLCs, sensors and other data sources. Cybernetics refers to this layer as the *hardware layer (1)*.

The first entity of the Boost 4.0 Big Data Pipeline is the *Integration Layer,* which facilitates access to data sources and infrastructure. In Cybernetics, this feature is implemented by the *machine abstraction layer (2)*.

The *data management and processing architecture* of the RA is implemented by the *message bus (3)* and *databases* for data persistence in Cybernetics. These components implement collection, linking, sharing and interactive batch processing of data.

The *data analytics* entity in the RA can be found in Cybernetics in various analysis components (4), digital twin plugins for Visual Components or predictive maintenance modules.

On top of this Cybernetics software stack there is the *business layer (5)*; Fill uses Cybernetics to implement customer specific smart factory solutions.

Finally, Cybernetics integrates machine and production data into its *data sharing platform (6)*, (c.f. RA), which enables advanced service development and engineering enhancement.

7.2.4 Integration of components

RISC and FILL are working on the scientifically close research project VPA 4.0 – Virtual Production Assistant funded by the Government of Upper Austria. The goal of this project is the conceptual design and development of a virtual production assistant (VPA 4.0) based on data and visual analytics of machine and job data, providing a learning-based knowledge base for the machine. The VPA 4.0 is primarily used to assist local experts. It allows a better understanding of the machine by recognising cause-effect correlations due to anomalies and patterns. In addition, maintenance intervals and cycles can be optimised and as a result, quality improvements of the production and the product can be achieved.

To achieve this, all data is imported into a **big data stack**, which is designed as follows:

Big data streams (for example sensor data of the production process or machine operation) are stored in the following way: At the backend, Apache NiFi [105] is being used to receive the online data, which is being stored into an HBase NoSQL data store [106], which is being accessed through Phoenix [107] for query speed. To enhance the flexibility and applicability of the system additional input data connectors will be implemented relying on standard importers and protocols such as OPC UA [108].

Model information of the machines is stored in the form of ontologies in a generic database system. The ontologies are generated by means of a graphical editor as well as further processing and additional data such as, for example, ERP data and machine configurations are stored in the generated data model by means of the ETL process.

RISC Software relies on the open source software collection *Apache Hadoop* for data storage and computation. Running on a cluster of computers, Apache Hadoop provides a software framework for distributed storage and processing of big data using the MapReduce programming model. Being originally developed for commodity hardware, it has also found use on clusters of higher-end hardware. All modules were designed assuming that hardware failures occur frequently and should be automatically handled by the framework, making it a robust and reliable solution.

The core component for storing data is the *Hadoop Distributed File System* (HDFS). Files are split into blocks and distributed across nodes in the cluster. HDFS achieves reliability by replicating the data blocks across multiple hosts, and thus RAID systems are not advisable for the data storage partitions.

To process data in parallel, Hadoop sends packaged code to each node and takes advantage of data locality, meaning that each node only manipulates the data to which it has access. This was found to be more efficient than conventional super computer architectures that rely on parallel file systems and distribution via high-speed networking. Beside the base modules, RISC software frequently uses Apache HBase, Apache Phoenix, Apache Hive, and Apache Spark.

HBase is an open-source, distributed NoSQL database running on top of HDFS. As a column-oriented key-value data store it can handle large quantities of sparse data. It is well-suited for fast read and write operations on large datasets with high throughput and low latency. It also features data compression and in-memory operation. Tables in HBase can serve as input and output for MapReduce jobs run in Hadoop.

Apache Phoenix is a SQL layer on top of HBase. It provides a JDBC driver that allows our customers to conveniently access and modify their data with various analytics and business intelligence applications via SQL.

Apache Hive provides SQL layers for various and file formats and file storage methods as well as NoSQL Databases that integrate with Hadoop, including Apache Phoenix. This means the unified SQL language HiveQL can be used to send SQL queries against a wide variety of data stores based on HDFS.

RISC Software relies on Apache Spark for cluster-computing. It provides high-level APIs in Java, Scala, Python and R and the machine learning library MLlib. Spark keeps track of the data that each of the operators produces, and enables applications to reliably store this data in memory. This is the key to Spark's performance, as it allows applications to avoid costly disk accesses.

Spark programs can be submitted from outside the cluster using Livy-Sessions, for example from within the web-based IPython interpreter Jupyter. This way, after the cluster has done some heavy lifting, results can be plotted immediately using python modules such as matplotlib. This allows for a rapid development of data analysis algorithms (for example for condition-monitoring, predictive maintenance, and so on).

FILL uses Visual Components simulation solutions during the different product life phases from the conceptualisation until the commissioning. By using simulation from the initial conceptualisation phase allows easily defining the initial configurations and calculating initial productivity parameters. The data can be reused in the engineering phase where more level of details in the virtual components are included and production data are validated. This allows FILL to test different configurations even before the real machine is built, saving costs, reducing delays, increasing productivity and efficiency.

Commissioning and ramp up phase, is one of the most critical phases as it is when the equipment starts its production life. The utilisation of simulation and virtual commissioning allows validating the code and operation in the digital twin. Connectivity interfaces, such as OPC UA, facilitates the validation of the code of robots and automation systems in the digital twin. As a result, the commissioning time is considerably reduced and the production ramp up is accelerated reducing waste of time, cutting errors, enhancing quality and improving productivity.

In addition, in BOOST scenario, the digital twin can validate the new production scenarios during the engineering phase and used it for analysing predictive maintenance case based on historical cases. The integration with the PLM will increase efficiency, as it will be possible to reuse the data available in the PLM and new modifications of the components easily updated.

7.3 VWAE real-time self-learning virtual factory 4.0

7.3.1 Short pilot overview

The **VWAE pilot** deals with improving the efficiency of the intralogistics operations using the latest digital technologies, which are being developed under the Industry 4.0 paradigm. The pilot has its grounds in the production plant located at Palmela, Portugal. Palmela plant has been increasing its size and production during the last years. The target in the pilot is to improve the intralogistics operations in the warehouse. An efficient integration and deployment of Big Data, ICT and automation technologies will help to increase efficiency and efficiently adapt to changing production requirements.

The scope of the VWAE pilot can be summarized in the following objectives:

1. **Financial benefits**

Logistics operations are always required in the manufacturing processes, since the parts arrive to the factory until the final product is ready to be delivered. The introduction of big data technologies, automation and ICT technologies, which improve efficiency in the intralogistics operations, maximizing the use of space, optimizing operations and avoiding operation errors, will reduce costs and drive to financial benefits.

2. **Flexibility**

Nowadays almost all the industries require facing fast changes in production, automotive industry is not an exception. In the pilot will be integrated the available technologies for adapting fast and flexible to the changing production demands driven by customer requests.

3. Ergonomics

Automotive industry requires the participation of human operators. Improve working tasks and automate repetitive, dangerous or low level boring operations will improve working conditions for the operators.

4. Human-Machine interoperation

The use of technology for supporting humans in the tasks, which cannot be automated, requires a new approach where the operators are comfortable and easily adapt to work with the new machines or technologies. Fast and successful integration of technology in the factory will lead to plant's efficiency and productivity.

5. Increase in the process efficiency

The automation and control of the process through a centralized database enables for a business intelligence approach to the warehouse system. Tools such as reporting and KPI's offer the opportunity to analyse and improve the system with real time data (on-line material flow such as unit load movements/sequencing and operation/fleet management).

To achieve the previously mentioned objectives the pilot focus in the warehouse operations from parts reception to commissioning. As one car requires large number of parts, relevant components are considered in the pilot. Automation and communication technologies together with Big Data analysis and support will drive the development of this pilot supported with the virtual factory platform, which will allow mirroring the current pilot in the virtual environment without interfering in the real plant.

7.3.2 Short component description

The proposed deployment for the VWAE pilot follows a digital twin approach (Figure 26). Relevant processes of the pilot plant will be mirrored in the virtual environment provided by Visual Components to build the digital twin. Once the digital twin has been validated with the data available from the real pilot new configurations can be tested and data generated analysed using UNINOVA Bid Data implementation. In order to facilitate the validation of the solutions developed without affecting the plant operations a cell layout with two AGVs from ASTI and the communication infrastructure from Telefonica has been commissioned.

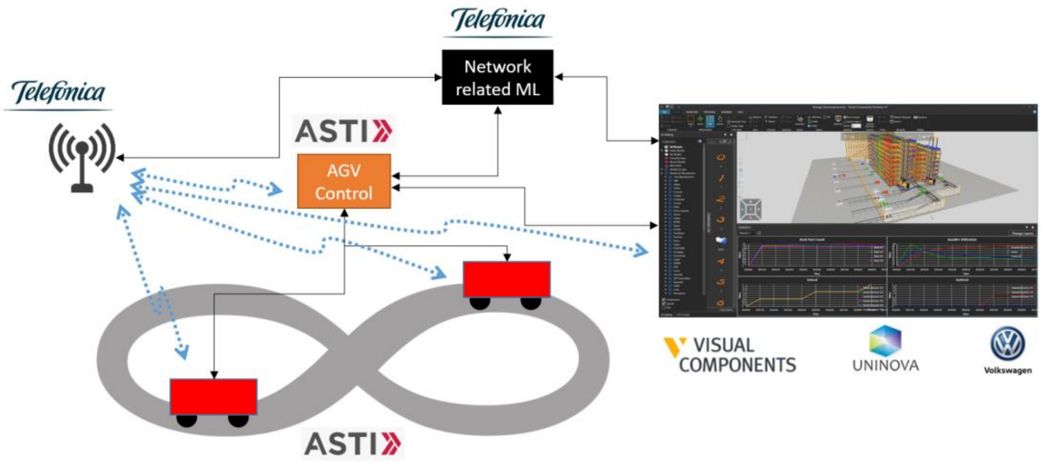

Figure 26 Picture describing the VWAE pilot and the different components.

ASTI will bring to the project two ASTI AGVs Easybot, one ASTI traffic control station A32, one ASTI charging station, some of magnetic tape used by the AGVs to follow a path, and some RFID tags to complete the design of one automatic intralogistics solution.

The **Easybot** is the best sold AGV in the ASTI's platform of AGVS. The main features of this vehicle are:

- Follow a track on the floor defined by a magnetic tape
- Execute actions according RIFD tag reading
- Control based on commercial PLC
- Circuit configuration can be done by the customer
- Speed: 18 m/min – 40 m/min
- Payload: 350 Kg
- Towing capacity: 250N
- Turning radius: 850 mm

The **A32** is one electronic board designed by ASTI to manage the traffic in the crossing points in the logistic circuits. It can manage until 4 independent traffic zones. The communication between the AGVs and the traffic station is managed wirelessly.

The **ASTI charging station** lets the AGVs to be working continuously, 24h 365 days per year. It is not necessary to take out the AGV from the circuit to start the charging process. The AGV comes equipped with brushes and one mechanism specially designed to plug to the charging station and receive the needed energy while it is working.

The **RFID tags** are a plastic-made discs with electronics inside. These discs are placed on the floor. The AGV comes equipped with a RIFD reader, the RFID reader periodically emits an electromagnetic signal. The tags enough closed to the reader receive this signal, which is used to feed the electronics inside of them. The tags also contain a permanent memory inside which is used to store information associated to the point in the logistic circuit. By using the energy previously sent by the RFID reader, this information is sent over the air and read by the RFID reader. This way the AGV can know in what part of the circuit it is.

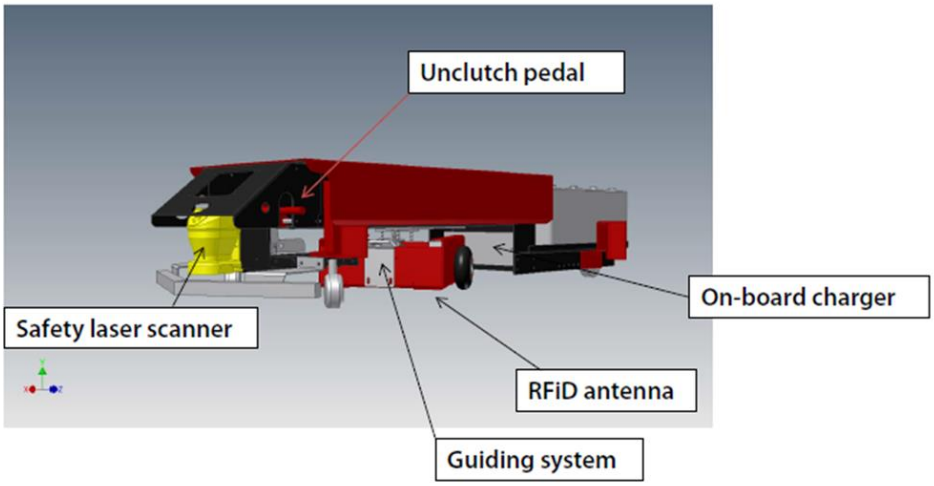

Figure 27 Description of the Easybot Robot

Telefónica I+D (TID) is the innovation company of the Telefónica Group. TID plans to contribute to the research activities of the pilot in two main actions:

- AGV communication improvement, via network data collection, normalization, big data analytics techniques and machine learning algorithms, applied over network traffic information. Using these inputs and techniques, impairments like faults, congestion, traffic loss, delays, etc., will be early detected, and feedback will be provided to the control components of the production environment. The aim is to reduce the impacts in close-to-real-time systems like that of the AGV communication.
- Validation over a 5G network: since at the time of execution of the final pilots it is very likely that 5G is still not widely deployed, TID offers the possibility to test and evaluate the associated solutions and technologies on the 5TONIC lab scenario, where last-generation mobile technology is continuously being evaluated and showcased.

UNINOVA will bring to the project an integrated big data integrated solution, which architecture and core technologies are depicted in Figure 28.

The general architecture and core technologies that support most of data processing work. Namely to efficiently gather, harmonize, store and apply analytic technics to data generated by sensors and other Cyber Physical Systems mounted on facilities and digital twin. The objective is to manage raw data, and from it, generate information to improve the supervision of all supply chain and production stages, from the shop floor to the office floor. Creating potential value from the data, by providing the tools to allow the development of new data- driven services and applications that can contribute to the objectives of zero default, near real-time reactivity to any problem, better traceability, and more predictability when manufacturing, meeting also the context of Industry 4.0 and Industrial Data Spaces. 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, being most of the technologies supported by the Apache Foundation. The architecture can be split in the following logic layers: Data ingestion layer, Data Storage layer, Data Processing layer, and Data Querying/ Analytics/Visualization layer.

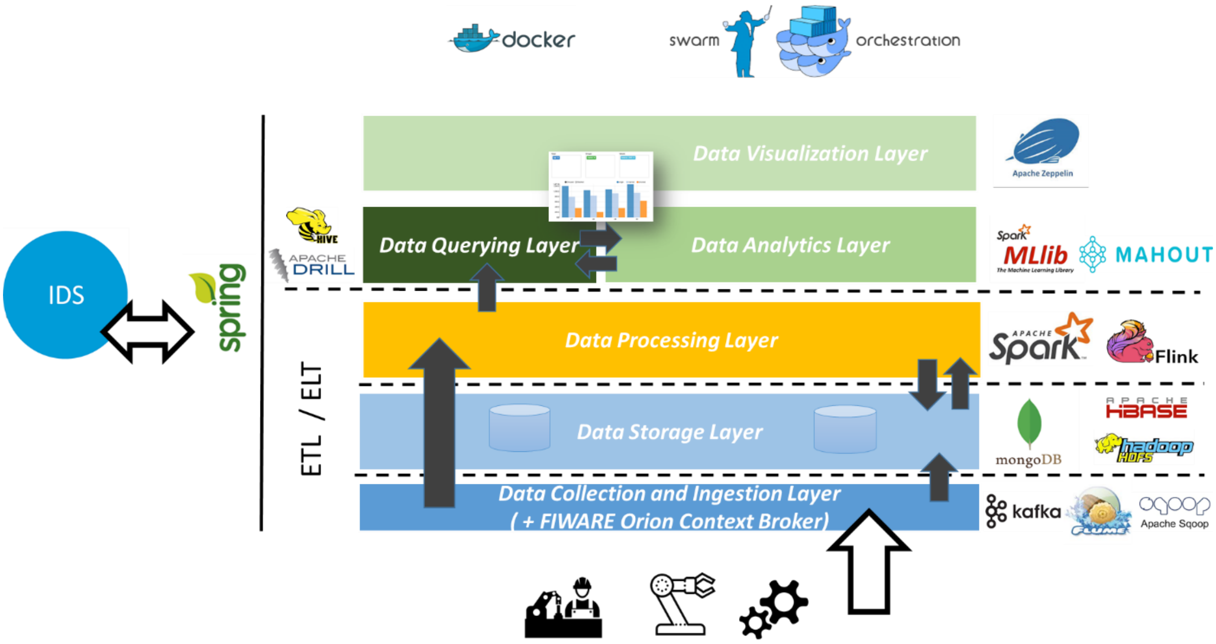

Figure 28 General Architecture and Core Technologies

Visual Components 4.0 Platform

Visual Components 4.0 is a 3D factory simulation and visualization platform to build any factory layout at different level of granularity. The platform includes on a set of innovative tools, which facilitates the different engineering processes, from conceptualization until commissioning.

The 3D simulation platform allows to create and visualize material flow, logistics and automation and robotic systems which can be programmed and validated, accelerating commissioning and reducing ramp up times. The open interfaces allow that solutions from different vendors can be validated simultaneously, saving time and reusing existing solutions.

Visual Components Premium 4.1 is the platform to be used in VWAE pilot. This solution (Figure 29) provides several functionalities to be used at VWAE Pilot:

- Easy to use and intuitive UI
- Large amount of pre-existing virtual models, parametric, component modelling.
- CAD import to create any new virtual component
- Layout configuration tools
- Large models capabilities
- Project ready deliverables
- Advanced automated and Robotics tools
- Communication interfaces, OPC UA and Web services.
- Open interfaces to interoperate with any other platform

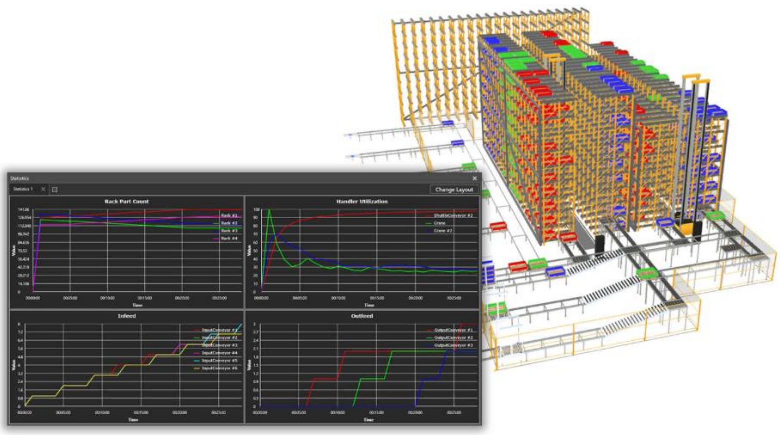

Figure 29 Screenshot of simulated automatic warehouse and performance analytics

7.3.3 Mapping to the RA

The solution for the VWAE pilot directly maps to the Boost 4.0 Reference Architecture, as shown in Figure 30.

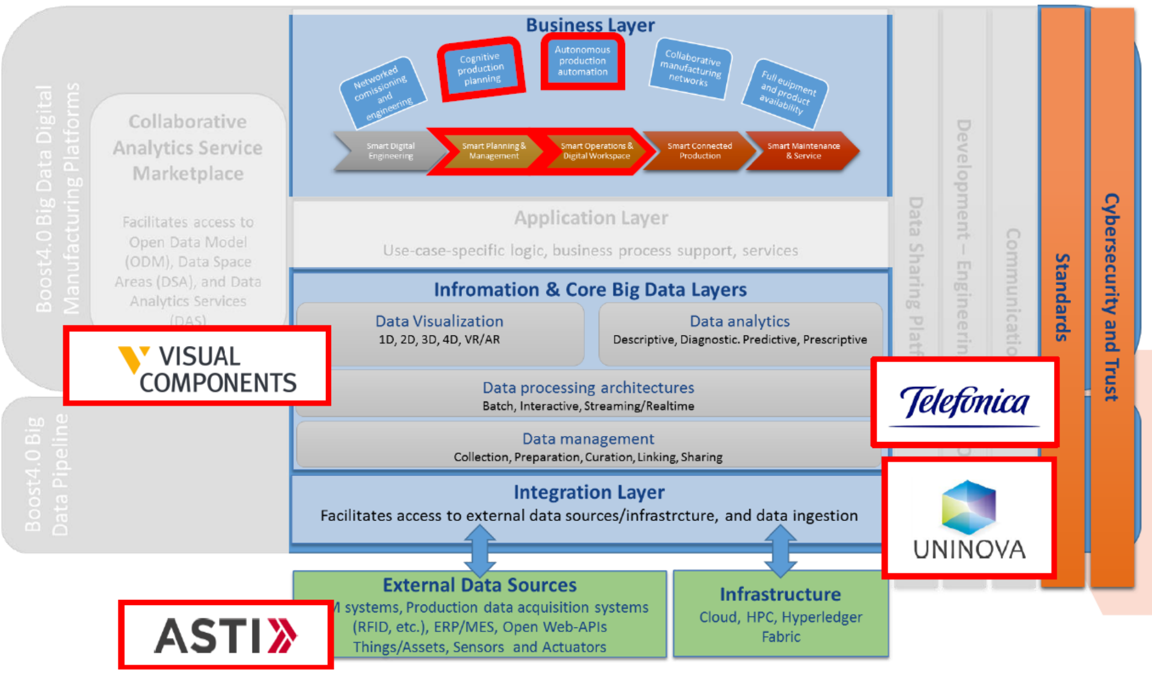

Figure 30 Mapping of the VWAE pilot to the Boost 4.0 Reference Architecture

On the bottom of the reference architecture are the physical devices, AGVs, sensors, actuators. In the pilot these are represented by ASTI equipment and the digital twin provided by Visual Components. The data provided from the equipment and digital twin is handled and processed though the second and third layers using communication and big data analysis technology.

7.3.4 Integration of components

The devised solution introduced in section 7.3.2 will be deployed in two different phases. In the first phase the virtual models will be created in the virtual environment and validated with historical data of the real plant and real AGVs. Once the virtual model has been validated it is possible that the data generated in the simulation can be used by the Big Data implementations, starting the second phase.

Bid Data generated in the simulation can be analysed to optimize the production environments. The utilization of the digital twin approach allows to easily create new production scenarios introducing new product variants, as well that will be possible to feed the digital twin with current production information from the factory floor.

In parallel, production scenarios obtained from the big data analysis can be verified in the real production cell, where it is available communication infrastructure and two AGVs.

7.4 +GF+

7.4.1 Short pilot overview

A critical issue for manufacturing in Switzerland for GF is the resulting machine cost with respect to main competition from other continents; differences are of the order of 30%, which can be partially compensated by the product quality and offer in terms of precision and productivity, but puts a high pressure on development, supply chain and operations. In particular it is necessary to improve the manufacturing and assembly of different machine functional components, groups (mechanical, electrical, electronic) and processes, but also be able to provide to the market innovative features in shorter times than the usual 2-3 years cycle for the upgrade or development of a new device. This involves a leap in data management and communication from the engineering stages towards the production stages until the usage phase, where a seamless feedback should be implemented in order to adjust the product quality, the machine manufacturing productivity and overall energy efficiency in real time.

In terms of assembly, even if some components supply are enabled to a just-in-time operation through automated communication of stocks, around 80 of machine component remain highly linked to human operations and databases which are not standardized across different production sites.

At the same time, it is crucial to improve the high level of precision and quality which are the backbone of the GF product differentiation. Traceability and monitoring tools are partially in place but they remain to be deployed across the lifecycle of the product, together with efficient predictive maintenance systems. Ultimately GF looks at the reduction of the total cost of ownership in order to provide competitive offers to the segments served, in particular aerospace, automotive, ICT and medical industries.

+GF+ machine tool optimum production factory 4.0 GF aims to create, digitize and standardize manufacturing related data across its value chain and beyond, to customers, and extract and process through artificial intelligence in a common data space the relevant information so to make it available to a smart production planning system targeting optimized, zero-defect manufacturing at a new GF factory in Switzerland.

7.4.2 Short component description

The platform should be open and have the capacity to merge heterogeneous data from milling machines and CMM machines. A system adapter in the open source platform will read all the data from milling and CMM machines. Relevant data should be identified in a pre-processing step using a visual analytics tool.

The data analytics infrastructure will be designed as shown in Figure 31. The continuous improvement loop (data / analytics / computation) will be performed to provide information for the workshop production monitoring that will help the decision making.

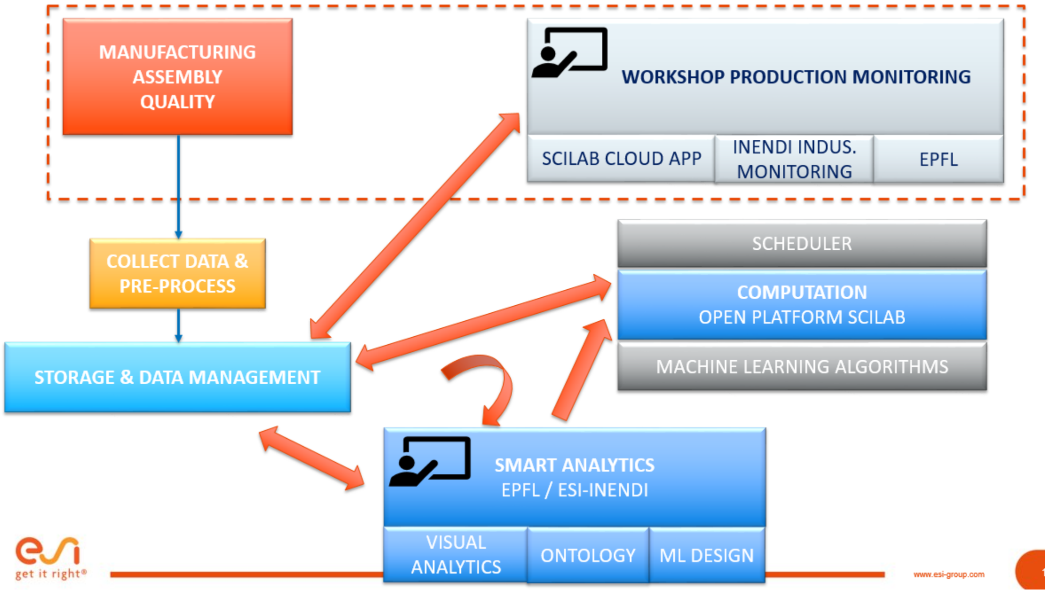

Figure 31 Data Analytics workflow

This coupled predictive-quality maintenance system will be able to provide more information than a standard predictive system based solely on machine sensor data, and will deliver correlations between quality/accuracy issues and potential component failure root causes.

For mold and die applications, the critical stage is the milling operation of the electrodes to be used for mold machining with EDM technology. Updating information of both the real geometry of the electrode and the strategies for subsequent operations will eliminate error propagation and accuracy degradation over time.

7.4.3 Mapping to the RA

The GF architecture maps to the reference architecture defined from the following points.

To address the concern of the acquisition and storage of data, application of the open source platform for the smart digital future considers **data management**. This open platform has the capacity of context management to merge heterogeneous data from **external data sources** (milling machines and CMM machines).

The **integration layer** in the open source platform will integrate all the data from milling and CMM machines. Tracking ID of a product produced by milling machine, it will recognize relevant measurement data. Afterwards, it will update attribute values of defined entities which have the context of the shop floor. Context data will be managed by Context Blocker within IDS connectors. Context data from the shop floor will be accessible through IDS connect which is a back-end processor communicating with the milling machine predictive maintenance system. Depending on data security and requirements, IDS connector will send relevant data to the **data visualization** and the **data analytics layers** for building the predictive maintenance system. The Predictive maintenance system will exploit distributed data for **the business layer** composed of maintenance planning & scheduling. This system is in charge of data analytics on operating data and monitoring, evaluation and prediction of the health status, and decision-making support. On the other hand, the open source platform allows exploitation of a specific part of milling machine data for the CMM precision management system which is a part of an operation management pilot.

7.4.4 Integration of components

The different elements (Figure 32) of the pilot integrate a seamless flow of data, from predictive maintenance and process monitoring of machines, to the real-time traceability of components in the assembly stage, to provide a framework for the development of a smart planning system leading to a zero defect manufacturing.

The GF pilot is aiming at developing an adaptive machining system in automated cells, integrated into an open cloud and data analytics infrastructure, taking in account the machine condition and performance. The platform will be able to diagnose in real time the state and performance of the machine and correct deviations through updated simulation models and planning systems.

The system should be automatically configurable and secure, to allow the seamless integration of all relevant data in a specific common control and analytics space.

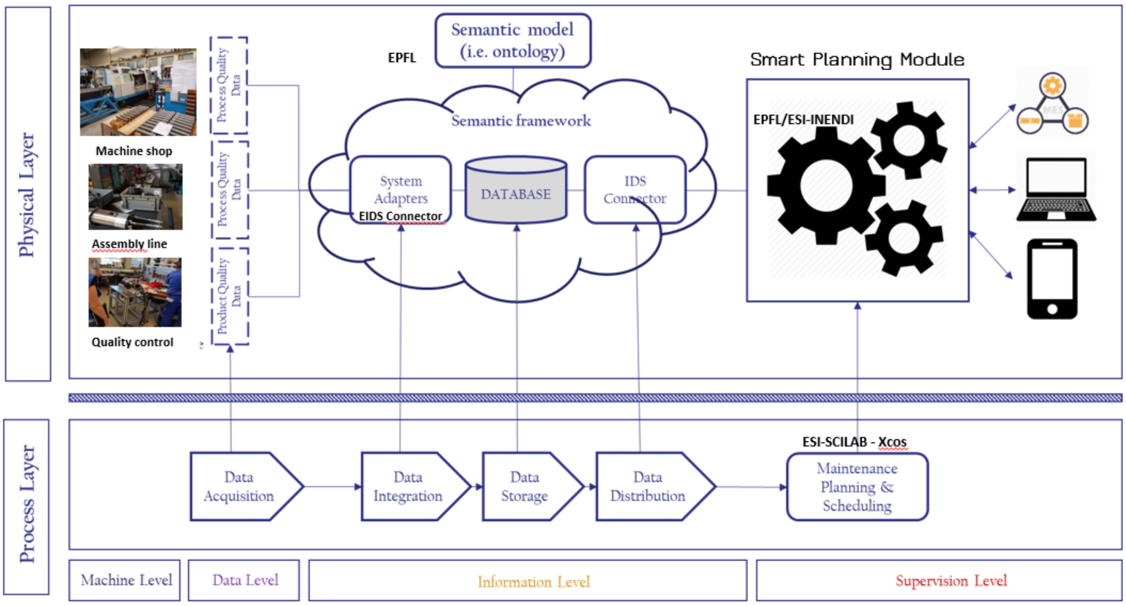

Figure 32 Smart Planning Framework

7.5 FIAT autonomous assembly line factory 4.0

7.5.1 Short pilot overview

The objective is the implementation of the concept of autonomous production, where the traditional linear process is removed and Mobile Robots, such as Automated Guided Vehicles (AGVs), Collaborative robots with vision capabilities collaborate with fixed production cells. Currently, mobile robots have only duties related to logistics (e.g. replenishment, preparation of components, etc.) or manufacturing (e.g. carrying work in progress) and the control of fleets of such AGVs and their availability and reliability to respect cycle time and lead-time is crucial to ensure the stability and throughput of the production systems.

The FIAT autonomous assembly line factory 4.0 aims to provide the maximum flexibility to potential changes in the demand or to issues/delays/changes in the logistics or productive systems by means of using available and new datasets (such as flows of components in the plants and their precise localization) ensuring business continuity, at the same time the over-dimensioned fleet of robots is reduced and the (big-) data are shared among the whole value chain (providers, maintenance services, etc.)

The trial will be focus on the following key areas:

- Connectivity
- Collaborating Platform

- Developing domain specific industry applications (and digital services)
- Enabling closed-loop innovation

The trial will demonstrate the quality and throughput increase through the monitoring and predictive maintenance of AGVs and production cells. In particular, the partners will focus on:

- Productivity improvement
- Flexibility
- Reduction of maintenance costs
- Reduction of operator cost
- Logistic management
- Quality management
- Achievement of zero-defect manufacturing

The solution will be demonstrated and applied on the FCA production site in Campus Melfi, which consist of assembly, welding, material handling, and quality control facilities.

The following relevant business processes are identified (more details in D2.1 Pilot Requirements and Use Cases Specification v1):

- BP 1 – Flexible and Smart Plant Operations through AGVs
- BP 2 – After sales service & Downtime reduction
- BP 3 – Analytics & Predictive maintenance
- BP 4 – Monitoring & Diagnosis

7.5.2 Short component description

In this trial, SIEMENS will provide an IoT Operating System (MindSphere), industrial know how and digitalization expertise. With Siemens global installed base of millions of devices (30 million automation systems, 70 million contracted smart meters, and 800 thousand connected products), Siemens can develop high-value applications through MindSphere's application programming interfaces (APIs) and deliver digital services with deep industry knowledge and experience.

MindSphere is the cloud-based, open IoT operating system that connects real things to the digital world, and enables powerful industry applications and digital services to drive business success. MindSphere's open Platform as a Service (PaaS) enables a rich partner ecosystem to develop and deliver new applications.

The solution will cover the following key areas:

- Connectivity
- Collaborating Platform
- Developing domain specific industry applications (and digital services)

- Enabling closed-loop innovation

CONNECTIVITY

Connectivity is one of the main topics in IoT: connecting devices is the first step of any digitalization project. The scope of the project will be to provide a standard technology to easily connect AGVs and the other devices involved into the project.

MindSphere allow to connect devices instantly with no programming skills or asset downtimes.

MindConnect Elements (software and hardware) take care of the security mechanism that connect and send data only to the MindSphere Platform.

During the on-boarding process, the MindConnect Elements must go through an authentication process with MindSphere. Once done, the two entities agree on cryptographic keys for use in further communications. Thus, the MindSphere Platform is designed to receive data only from valid MindConnect Elements, which have successfully completed the authentication procedures during the on-boarding process.

As the level of digitalization increases, so too does the importance of comprehensive security concepts for applications. With defence in depth, Siemens provides a multi-layer concept on security, network security and system integrity as recommended by ISA 99/IEC 62443 and IT Security oriented to industry standard ISO 27001/BSI. The data in motion is always at least 256 bit SSL/TLS encrypted.

All communications between the MindConnect Elements and the MindSphere Platform are encrypted via the Transport Layer Security (TLS) 1.2 standard. The TLS configuration is regularly checked to comply with the applicable Siemens Information Security guidelines. This helps in protecting against man-in-the-middle attacks or any manipulation of communication to the MindSphere Platform.

COLLABORATING PLATFORM

MindSphere open platform supports developing applications. MindSphere has several APIs at developers' disposal to enable development of customer owned applications. These APIs are optimized for industrial IoT application development thanks to additional reusable supporting modules (e.g. for visualization, analytics and parsing).

The scope of the platform is to facilitate the implementation of the domain know how (i.e. predictive maintenance algorithms) into a MindApp that can be used by any end user.

DOMAIN SPECIFIC APPLICATION

MindSphere provides a solid foundation for applications and databased services. Siemens, or any other partner with development skills, will implement an application for the predictive maintenance.

CLOSED-LOOP INNOVATION

Siemens Software and Digital Services enable digitalization of the entire lifecycle, seamlessly integrating product, production and service processes. MindSphere connects to real-world devices and processes, and feeds performance data back into a performance digital twin to drive improved decision-making and intelligence.

The architecture of the platform will be deployed in the pilot as described in the following diagram:

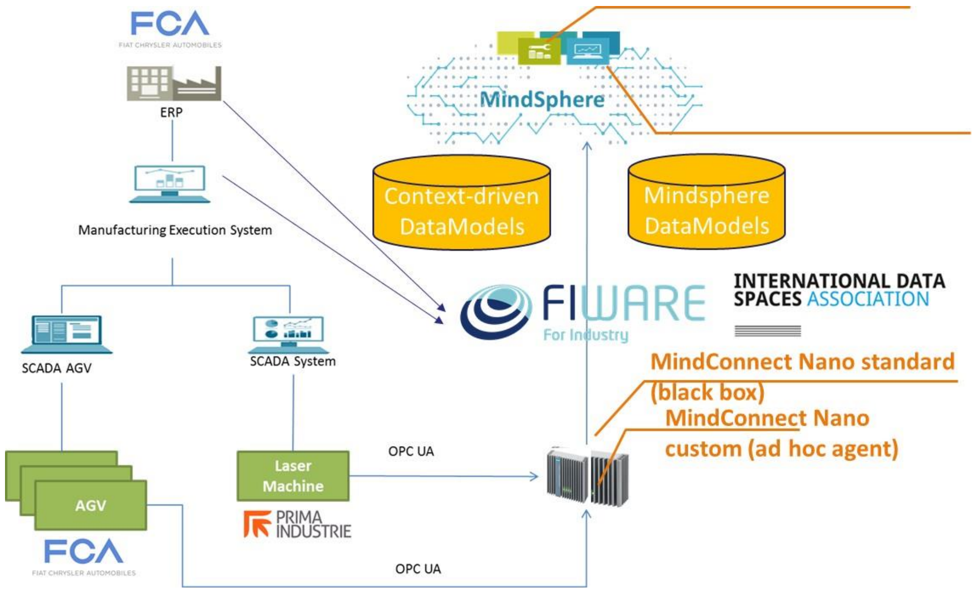

Figure 33 - FCA pilot architecture

7.5.3 Mapping to the RA

The FCA pilot focuses on the development of a Big Data infrastructure for the Monitoring and Predictive diagnostics of AGVs and Production cells in an Autonomous shop-floor. The scenario architecture reflects the Boost 4.0 reference architecture as shown in Figure 34.

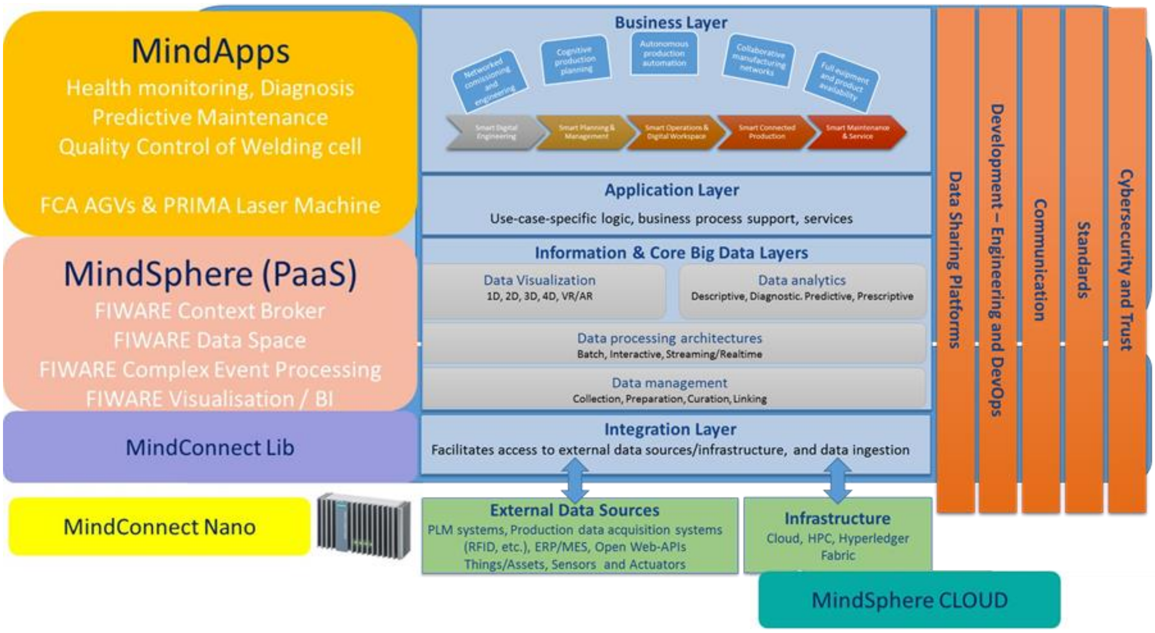

Figure 34 - Reference architecture implemented for the FCA pilot

Datasources coming from the shop-floor, in this scenario, include the vibration and other mission's data (AGVs) and quality-related data (PRIMA cell). On top of these, MES (Manufacturing Execution Systems) and WMS (Warehouse Management Systems) data will be used to correlate sensor data with production planning and control data. Preliminary analysis (outlier detection, data verification and preparation) is performed at EDGE level. Then data are provided through the MindConnect Nano or other BOOST 4.0 adequate hardware and protocols. The MindConnect libraries are used as integration layer towards the cloud.

Regarding the Data management, Data processing, Data visualisation and Data analytics, several FIWARE tools will be used and exactly: the FIWARE Context Broker, Data Space, Visualisation Business Intelligence (BI) and the CEP (Complex Event Processing).

It is crucial that the data broadcast to the cloud is adequately encrypted, together with customised rules for data sharing enabled by IDS.

7.5.4 Integration of components

All the components will be integrated together in the integration layer of the RA model. Sensors data coming from the machines will be pre-processed and sent to the platform. Simple EDGE computing will be performed, while the (Big) data will be sent, analysed and displayed through the integration of FI-WARE and MindSphere components. Using IDS data format will ensuring that key industrial data will be adequately protected.

7.6 PHILIPS Autonomous short-batch injection moulding production process

7.6.1 Short pilot overview

The pilot focuses on using data from injection moulding machines available at PCL in Drachten. Today there are different models and versions of the machines. This leads to different types of the collected data as the machines use different communication protocols, data structure and also different data features. Also, the data from most of the machines in not accessible for real-time use. A consequence of this is that all data related projects are set up individually by the Big Data Analysis team and they are based on the requirements needed from the injection moulding team that are in need of further insights.

This pilot aims to standardise the setup for how all the collected data should be treated and managed. This has mainly to do with the requirements of the injection moulding line, to increase available data and the availability of the Big Data Analysis team. The preferred approach would be to have a platform, which would facilitate a "plug & play" module for all injection moulding machines present at PCL in Drachten, automating the steps of:

1. Managing the data storage,
2. Deployment of models,
3. Management of models,
4. Scoring of models.

7.6.2 Short component description

A modern extendible platform for storing and managing raw and pre-processed data, for hosting micro-services and data models and also that enables access to historic and real-time streaming data. Such an integrated component does not come out-of-the-box, but can be achieved by combining existing open-source technologies such as the Apache Kafka ecosystem, Kubernetes, Apache Spark combined with notebooks and the many available data store technologies.

A data pre-processing interface that will: (1) receive raw data from all the different moulding machines, (2) pre-process the data to fit standardised data form and (3) send the pre-processed data to a platform.

Data analytics models that will be deployed in the pilot will mainly focus to satisfy the following two strategic goals setup in the pilot:

- Predictive Quality – improvement on Fall Off Rate
- Predictive Maintenance – less down time

7.6.3 Mapping to the RA

Information and Core Big Data layer: The platform will have the following functionalities:

1. Data management: the platform will collect and store the raw data and manage the transformation of the raw data into the standardised form. The platform will also share the data for pre-processing and with data models.
2. Data processing architectures: The platform will support different modules for pre-processing the raw data. This to better serve the data analytics layer. We will look into both batch and stream analysis data flows. Data clean-up is also something that will be introduced in this layer. A gradual approach will be adopted and an effort will be made to model the findings from the initial experiments.
3. Data analytics: the platform will host data models for predictive maintenance.
4. Data visualisation: share the insights with the dashboard

Application layer: data models will be tailored to the needs of this pilot. They will focus on predicting the faulty parts of the moulding machines.

7.6.4 Integration of components

The moulding machines will send the raw data. The data will be stored on the platform, where it will also be pre-processed to fit the standardised form. The data in this format will be stored again on the platform. For the standardised format a canonical extension of the IDS and Fiware data formats is considered.

Data models will be deployed either on edge or on the platform. They will access the data on the platform. The results will be visualised on a dashboard.

7.7 GESTAMP automotive part prescriptive quality assurance factory 4.0

7.7.1 Short pilot overview

Gestamp is an international group dedicated to the design, development and manufacture of metal automotive components. The Group specialises in developing innovatively designed products to achieve increasingly safer and lighter vehicles, thereby reducing energy consumption and environmental impact. Gestamp has identified Industry 4.0 as one of the main global objectives.

Nowadays this global phenomenon, Industry 4.0 intends to "give intelligence to things, processes and systems through data analysis". Gestamp is fully committed with Industry 4.0 and the well-known concept of Smart factory, creating more efficient manufacturing plants and more consistent and reliable processes through the analysis of data, by adding intelligence to the processes and getting the right information to the right people. In short, Gestamp pursues to increase the supply chain, in terms of process efficiency to meet the global needs of their customers in quality, time and cost.

The group has developed strategic initiatives to achieve Smart plants concept, across 4 principal pillars:

- Energy Efficiency: Reduce consumption and reduce oversized equipment.
- Maintenance: Reduce corrective maintenance and increase efficiency.
- Logistics: Optimise internal logistic functions through task prioritisation.
- Quality: Focus on zero defects to customers.

In spirit of this, Gestamp pilot will focus on two of these four main pillars in order to improve the overall efficiency of the plants, by extracting and smartly organising the present and future big data acquisition that somehow, they already generate and will generate in the future; making it accessible so that enhanced visualisation and (predictive) analytics can better support the decision-making.

1. **Real-time Indoor Assets location platform optimisation and secure data**
   This business scenario will be orientated on the logistics efficiency, through the real-time indoor location solution optimisation, Gestamp will be focused on:
   - Operational effectiveness.
   - Increase flexibility of production.
   - More dynamic allocation of resources and capacity definition.
   - Reduce changeover time.
2. **Zero Defect Manufacturing powered by massive metrology**
   The general goal of this business scenario is to improve the performance and efficiency of the quality control process to achieve zero defect manufacturing. The aim is to improve current metrology process by implementing an innovative high-definition metrology process, gathering, integrating and analysing quality data from different sources and data from the product design manufacturing step, and develop a collaborative and visualisation implementation as well as advanced analytics. Moreover, high-density processing and visualisations process will be

developed in order to ease the analysis of massive point clouds, coming from complex products which will allow a deeper knowledge about components and products. Hence, to reach this goal, the future scenario is aimed at developing a more efficient quality control process, covering the whole quality control workflow based on Quality Information Framework (QIF) standard.

In summary, the expected outcomes to achieve resource efficient factory are:

- **Quality customer satisfaction impact**
  Knowledge from individual components up to complex ones will allow to develop improvement action plans to reach zero-part defects.
- **Process performance improvement**
  Through real-time assets allocation, internal process will be constantly managed becoming more flexible, efficient and cost-effective.
- **Supply chain collaboration impact**
  Combining real-time assets data with MES & ERP systems data will provide profits in terms of warehouse and storage area management, on-time deliveries and inventory processes.

7.7.2 Short component description

7.7.2.1 Real-time Indoor Assets location platform optimisation and secure data

The Real-time Indoor Assets location platform optimisation and secure data scenario will be developed based on ENEO's redborder platform and I2CAT's solution.

ENEO will use its redborder platform as the basis for its contributions to the business processes contemplated in Business Scenario 1 of the Gestamp Trial. Specifically, redborder will be used to provide advanced data visualisation for asset location tracking to achieve smart logistic optimisation as well as provide detailed analysis of related production data, resulting in the creation of dynamic user dashboards. Additionally, the redborder platform will provide an added layer of security to ensure that the communications between connected factory assets is secure as well as a framework for the detection of anomalies in all of the processes analysed.

The redborder Network Traffic Analysis platform provides the framework upon which different modules offer a number of unique functionalities. For the GESTAMP Trial, the following modules will be used and adapted to the specific needs of the business scenarios and processes.

- Traffic: redborder's Traffic probes are compatible with the leading industry standards for network traffic data including Netflow v5/ v9, IPFIX and Flexible Netflow protocols - sFlow and jFlow. The Traffic module offers a simple and straightforward method of gathering knowledge about the flow of traffic on any network.
- Vault: the Vault module is a SIEM system which can analyse and manage a large number of logs, offering the possibility of detecting more complex trends in the data analysed. The Vault module provides functionalities such as metadata extraction, data enrichment, correlation, and storage, all of which will be utilised in the trial.
- Mobility: The Mobility module allows WiFi and location information to be captured for connected devices. In the trial, this module will be used to provide precise location based on WiFi connection for the connected factory assets.
- IPS: The IPS module allows users to centrally configure, supervise, and apply security policies in the open source realm. It is hierarchical, multi-tenant and multi domain structures control thousands of devices. The results of the real-time analysis are presented in easy to understand dashboards to provide administrators with the information necessary to take action. Automatic responses to events which meet predefine criteria are also configurable through the redborder manager.

The architecture of the redborder platform to be deployed in the pilot is presented in the following diagram:

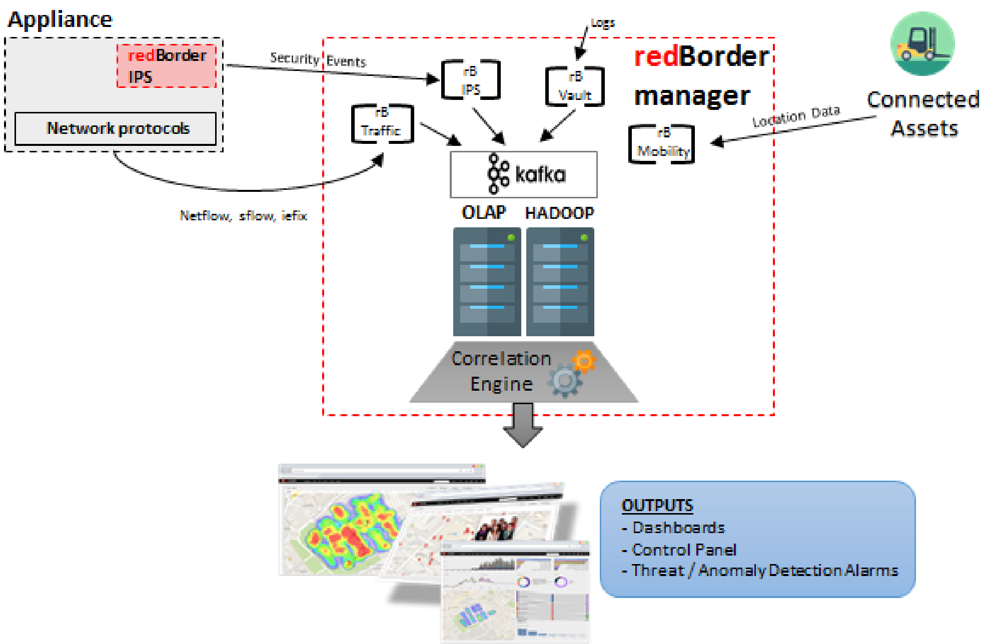

Figure 35 redborder architecture

I2CAT tracking solution main goal is to provide a dedicated RTLS (Real Time Location System) based asset and product optimisation and its implementation on Industrial environments as well as data acquisition modification: Define optimal solution based on all Indoor locating technologies (UWB & RFID) and considering different area requirements (Accuracy, required data and axis location – xyz). Furthermore, with data integration from the different sources of data, on a real-time basis, Gestamp has the possibility to optimise resources and facilitate flexibility in terms of assets routing.

With the Assistance on the data analysis, on the visualisation and the interconnection with existing interfaces (MES & ER) a shop floor performance improvement activity becomes real and on real time through secure data and assets management. The components of this solution are:

- **Ultra Wide Band (UWB) Positioning** is responsible for calculating the relative positions of each UWB tag against multiple UWB anchors. Using triangulation, the system can locate an asset inside the factory. UWB Positioning runs in any Linux based Operating system. UWB Positioning allow other applications to get each device location thanks to a pub/sub interface. It also contains an input file in which the different parameters of the UWB anchors can be changed over the air, without network collapse.
- **I2Tracking** is a unique software developed by i2CAT. It is designed to provide seamless integration between positioning technologies like UWB, BLE, GPS… and RFID readings, matching each event to a specific location and determining the different states of the objects that are being moved around different areas.
- **Scalable Industrial data** is a distributed data solution that creates a modular pipeline of data from its acquisition to its external sharing. The data can be obtained at several points in the chain between the collection and the external distribution (sharing). Using this approach, a data consumer (application) can be seamless moved around the chain and placed closer to the data collection or closer to the data sharing, depending on the requirements.

7.7.2.2 Zero Defect Manufacturing powered by massive metrology

The Zero-Defect Manufacturing (ZDM) powered by massive metrology scenario will developed based on TRIMEK's Big Data M3 platform and Capvidia MBD/MBE solutions.

The M3 platform is poised to provide a structured solution for Metrology4.0, an edge-powered quality control analytics, monitoring and simulation system. This solution is used for the organisation, analysis and reporting operations of the metrological information, taking advantage of the storage and computational capabilities of the cloud to carry out advanced operations and provide smart added value services. Figure 36 depicts the M3 global architecture.

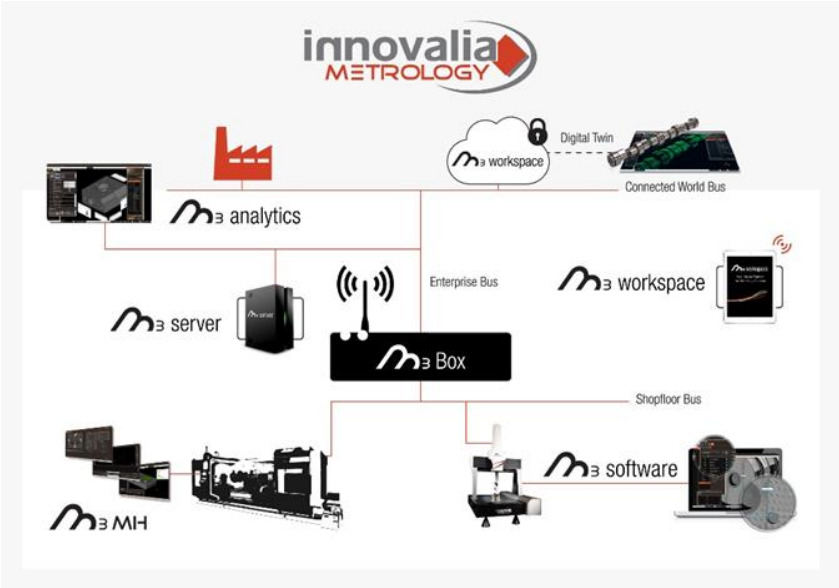

Figure 36 M3 architecture

The ZDM powered by massive metrology scenario is focused on improving the efficiency of the quality control workflow by implementing innovative technologies like high-density metrology or virtualising massive metrology in order to achieve zero defect manufacturing. Thus, this scenario consists of:

A. High-density metrology for high-performance processing of massive point clouds and advanced multi-level visualisation with realistic colourmaps with textures for complex products like chassis.
B. Virtualised massive metrology for the implementation of a complete QIF metrology workflow, starting from the product design process up to the advanced analysis and visualisation of the quality information, ensuring the interoperability of the manufacturing information and the traceability of the manufactured parts and products.

Considering M3 platform, main components to be used to succeed in achieving ZDM are:

- M3 software: The M3 software is a high-performance software for capturing and analysing point clouds in a versatile, agile and powerful way. It covers the entire spectrum of metrology. It works locally but is powered up by the use of the edge-powered technologies included in the global solution. The aim of this component within the pilot will be the rapid acquiring and processing of massive point clouds and multi-level visualisation of complex products, as part of a unique QIF metrology workflow (Quality Information Framework standard).

- M3 Workspace: It permits the massive management of digital parts and point clouds, storing and sharing the metrological information. It acts as a sort of repository where all the metrology results are located and allows its further analysis and visualisation.
- M3 Analytics: The M3 analytics is a powerful tool that enables the advanced analysis, its visualisation, and the reporting operations related to the massive metrology data stored in the cloud by means of several algorithms and computational components.

7.7.3 Mapping to the RA

Gestamp pilot focuses on Real-time Indoor Assets location and secure data and Zero-Defect Manufacturing powered by massive metrology in order to gain knowledge about products and processes, achieving eventually a collaborative manufacturing network to enhance the efficiency of the manufacturing plants.

7.7.3.1 Real-time Indoor Assets location platform optimisation and secure data

The architecture of this scenario is completely aligned with the Boost 4.0 architecture as shown in Figure 2. The different data sources of this scenario, namely the Container tags (RFID), the UWB distance to every antenna, the acceleration and the data from the enterprise information systems will be integrated in the Cloud. Outlier detection will be performed in the Data management layer to assure the quality of the gathered data and to discard malfunctioning or wrong measures. Data processing techniques will be used to perform Big Data analysis support either in batch (Historical) or stream (real-time). Data visualisation will consist on simple in plant visualisation for the technical staff to locate the asset. Applications considered will be part of the MES of the Industry 4.0 so that Business questions can be asked (and responded).

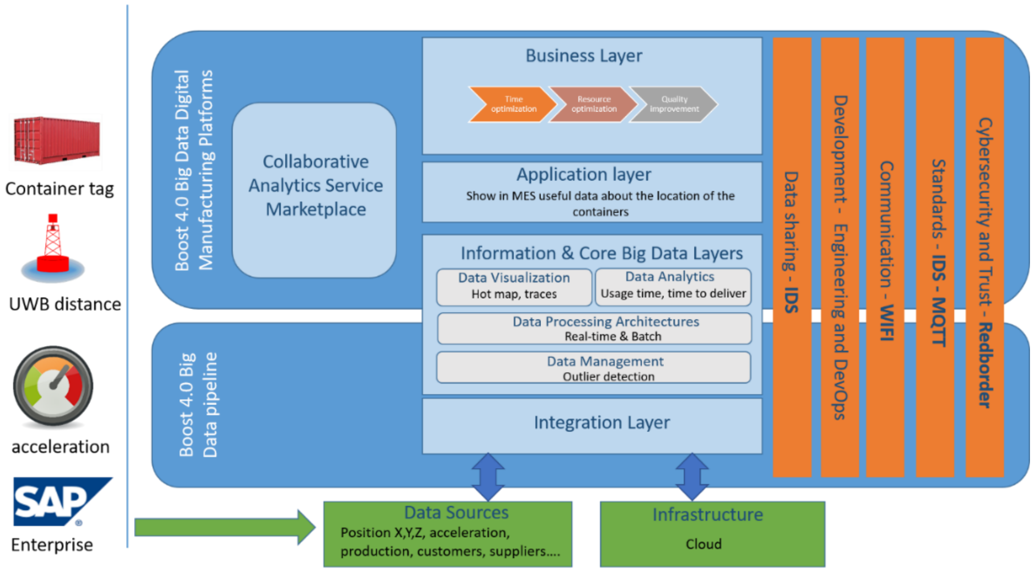

Figure 37 Gestamp pilot data sources

All the transmissions will be encrypted end-to-end and moreover, standard data formats will be used like IDS or MQTT. The communication channel will be industrial WIFI while the data sharing will be done though IDS.

Moreover, the redborder platform provides advanced IDS / IPS capabilities based on the establishment of customised rules for each installation. Therefore, it is well aligned with the security and trust layer. The main goal of this scenario is to ensure that the factory network is secure and that the data can be used throughout the operation without fear of loss or compromise. Additionally, the platform provides industry-leading data analytics and visualisation capabilities. These capabilities, once proven in the trial setting, will allow users to understand and manipulate the data generated by the factory assets and network connected devices in innovative and novel ways. Decisions will be made based on more accurate, trustworthy and secure data.

7.7.3.2 Zero Defect Manufacturing powered by massive metrology

The high-density and virtual massive metrology permits an accurate control process based on QIF standard with innovative and efficient visualisation, processing, storage and analysis capabilities, which will definitely improve the decision-making process and reduce the defective parts.

BOOST4.0 solution must deal with data gathering from inline Coordinate Measuring Machines (CMMs), product design, IIoT, 3D scanners, tactile probes and so on, and further data processing, visualisation and analysis making use of cloud infrastructures. Acquired data, like massive point clouds will be managed, curated and shared for real time processing and/or batch processing when required. The processed metrology data will be available for high-performance visualisation (for example realistic colourmapping with textures) and advanced analytics such as statistics, time-series patterns… for diagnosis and prediction. These services will allow to ease the decision-making process and increase the efficiency of the plant, having a smart connected production and smart digital engineering.

All the layers will use cybersecurity and trust technologies, standards like Quality Information Framework (QIF) and data shared among different platforms to ensure security, trust, interoperability, flexibility and traceability.

Figure 38 shows the BOOST4.0 RA with highlighted components which are the ones to be used within this scenario.

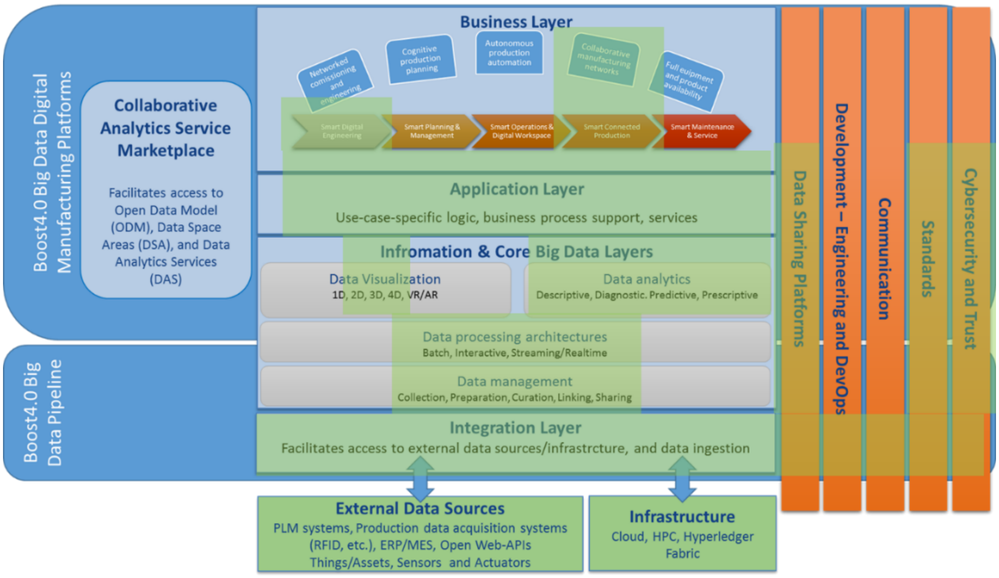

Figure 38 Gestamp pilot mapped in BOOST4.0 RA.

7.7.4 Integration of components

All the components will be integrated together in the integration layer of the RA model.

7.7.4.1 Real-time Indoor Assets location platform optimisation and secure data

By providing the tools necessary to understand and analyse the position and activity of all factory assets in real time, as well as monitor all network activity, the proposed solution will have an immediate and measurable impact on the factory's decision-making process. The ability to monitor network and asset activity in real time will allow decisions to be made which take into account more accurate data and lead to more efficient outcomes.

<u>IDS and Fiware</u>

This scenario will be integrated with IDS module provided by FIWARE to share all the gathered information with third parties that can provide new data analysis or business analysis solutions to the plant. Hence the IDS adapter will be deployed right after the RabbitMQ [109] (MQTT transmitting queue), being a consumer of the messages produced by the asset. To completely isolate the production environment from the external access, a middleware will be used based on WSO2 [110].

Big Data Europe components

Two main components of the Big Data Europe project will be used in different layers of the Architecture, namely:

- Spark: Stream (real-time) processing of data.
- Hadoop: Batch (historical) processing of data.

In this scenario, Kafka will not be used as data broker since the low-level Industry based environment where this scenario takes place needs solutions more adapted to the GESTAMP plants. In this case RabbitMQ will be used following the known and adopted standard of MQTT for data transmission.

7.7.4.2 Zero Defect Manufacturing powered by massive metrology

This scenario deals with future metrology which will be used to assess and guarantee the fit, performance and functionality of every part. In spirit of achieving the goal of ZDM powered by massive metrology, the following components have been identified as essential part of BOOST4.0 platform for Gestamp pilot:

IDS and Fiware

The scenario will use on the IDS (IDS Connector) and FIWARE (FIWARE Context Broker) technologies in order to share massive metrology information and product design information to be used for the advanced analytics and visualisation and/or external smart apps. In this context, the interoperability within the metrology workflow as well as to other platforms can be guaranteed.

In particular, FIWARE-based IDS Connectors and the corresponding system adapters will be deployed in order to preserve a secure and trustworthy exchange of metrology and product data between data providers (TRIMEK and Gestamp) and consumers (TRIMEK or external apps) in separate systems, where the provider has the full control over the data that is being shared with the consumer. Using this

Big Data Europe components

Moreover, the high-density and virtual massive metrology will leverage on some BDE components for data management and data processing as it is depicted in Figure 39.

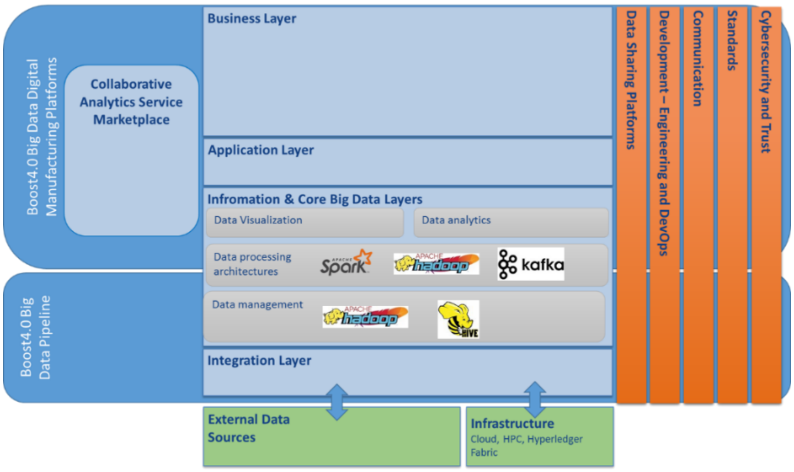

Figure 39 BDE components for high-density and virtual massive metrology

7.8 Volvo truck digital assembly factory 4.0

7.8.1 Short pilot overview

The aim of the implementation of the trial from the end user perspective is to address challenges (delays, and re-ordering to racks) in the supply chain between two production plants. A solution is required that can track the components (truck-cabs in this trial) from the cab production plant located in Umeå to the final assembly plant located in Tuve, Gothenburg. The components are transported on racks, so another objective is also to enable tracking of the racks when returning back to the Umeå plant. Besides tracking and providing real-time information of the location of cabs and racks, the data that is collected over time is expected to provide insights about the trends in delays of cab and rack transportation.

The proposed trial involves the cab production plant in Umeå and the final assembly of tracks located in Tuve, Gothenburg. The trial focuses on:

- Allowing full visibility of the supply chain.
- Forecasting of the arrival time of the cabs/racks and to detect delays in the flow.
- Forecasting the arrival time of racks in the return flows to Umeå.
- Optimisation of assembly planning based on the live monitoring and forecasting of the supply chain.

7.8.2 Short component description

7.8.2.1 Data analytics tool

The Big Data Analytics tool is a tool that enables the forecasting of cabs arrival time and enhance the decision making over the supply chain. Data are provided from a track and trace mechanism for the transportation of cabs and racks in the VOLVO supply chain. The available data are queried from the blockchain and a big data storage. In addition, the analytics tool takes into consideration many factors based on a vast amount of historical data coming from VOLVO applying statistical and machine learning methods in order to provide the highest possible level of accuracy of the forecasting estimations.

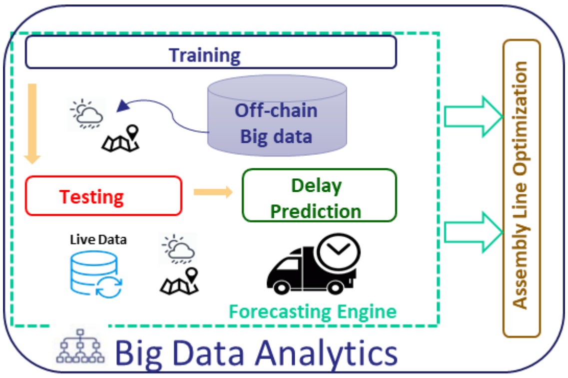

Figure 40 Big Data Analytics tool architecture overview.

In accordance with Figure 40 the Big Data Analytics tool comprises the following phases:

- Training phase:
    - **Route Modelling** is the process of segmenting the Umea to Tuve path in smaller trajectories to achieve a better analysis and detection of the possible transportation delays. Distinct trajectories are created based on the transportation route. The separate trajectories have to be matched to the corresponding road segment. **Map matching** algorithms map each GPS point of a trajectory to the actual road segment and the GPS data are translated into manageable information.
    - The **Feature selection** process is about choosing the appropriate factors that the arrival time of the cabs depends upon. These are:
        - the lack of material in Cab production in Umeå
        - lack of racks
        - machine breakdown
        - product damage during transportation
        - weather conditions

    - traffic
  - The **training of historical data** is the implementation of state-of-the-art methodologies for the prediction model construction. These are:
    - families of Artificial Neural Networks (ANNs) [112],
    - Support Vector Machines (SMVs) [113],
    - Classification and Regression Trees (CART) (decision trees) [114] and their ensemble approach (Random Forest) [115],
    - Markov chain (MC) [116] and Hidden Markov Models (HMMs) [117].
- **Testing Phase and Delay Prediction:** In each part of the route data analytics methods are implemented, taking into consideration several factors that could affect the transportation time. The training model will be optimized on the testing phase by evaluating the model and take into consideration the accuracy and precision of the prediction models. Delays are detected in each road segment and eventually lead to the overall delay prediction in the transportation flow. Thus, the prediction model will operate and delays will be detected in each road segment and eventually lead to the overall delay prediction in the transportation flow.
- **Assembly Line Optimization:** The final component of the data analytics tool is the assembly line optimization mechanism that provides information to assembly line planning departments for replanning and re-optimisation of cab trim and final assembly at the final assembly plant. Genetic algorithms and Constraint Programming are used for solving the assembly line optimization problem.

The data analytics tool is developed on CERTH IoT Platform. In the context of the VOLVO pilot dedicated to assembly line visualization and forecasting over the supply chain, the IoT Cloud platform has to be extended in order to fulfil business scenarios presented by the pilot and cover the needs of data storage, maps visualization and predictive algorithms and become compatible with EIDS and IBM Hyperledger Fabric.

More details about analytic algorithms and IoT platform are available at D3.3 Big Data Models and Analytics Platform.

7.8.3 Mapping to the RA

CERTH Data analytics tool will support predictive analytics models and techniques that will enable the forecasting of cabs arrival time and will enhance the decision making over the supply chain. The tool will build upon a CERTH IoT platform able to support Big Data Storage and analytic functionalities. The predictive analytics tool will be based especially on logged data for GPS sensors that will be available to the IoT platform. In Figure 41, the mapping of Big Data Analytics tool to BOOST 4.0 RA is depicted. Concretely, the Forecasting Engine and the Assembly Line Optimisation components component correspond to the

Data Analytics of the Information & Core Big Data Layers of the RA. Likewise, the Data processing architecture is entailed in the Forecasting Engine where data are retrieved and processed. The integration of CERTH IoT platform correlate to the Data management layer. Finally, the External Data Sources are Weather APIs, VOLVO ERP and the set of GPS sensors that will be installed.

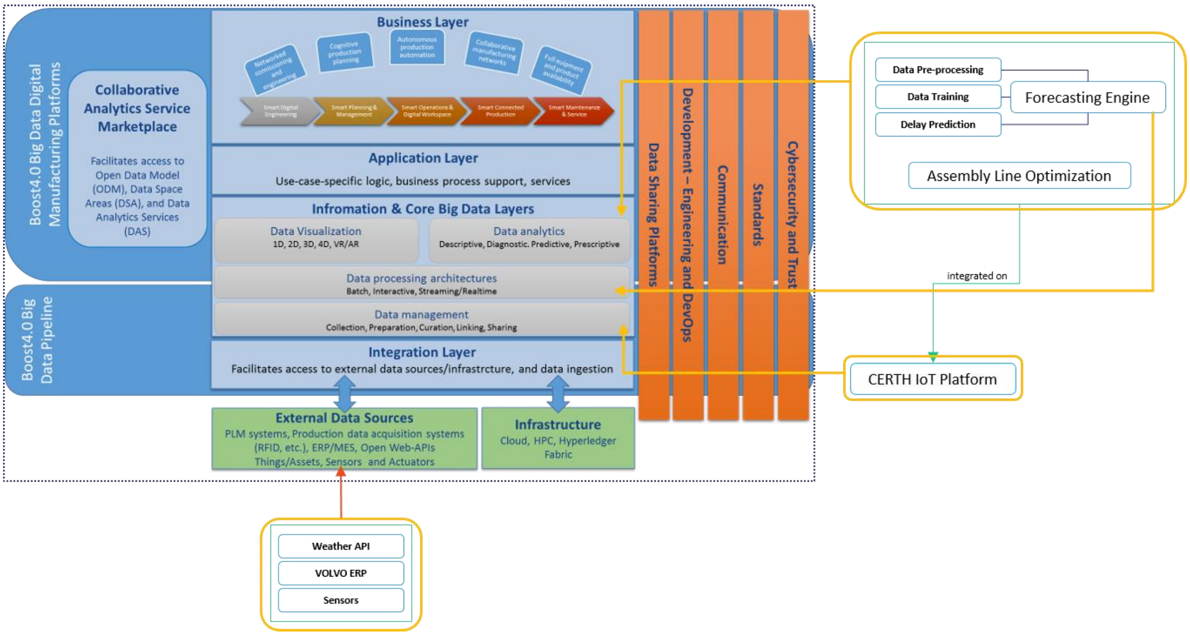

Figure 41 Mapping of Big Data Analytics Tool to BOOST 4.0 RA

The Hyperledger Fabric mapping to BOOST 4.0 RA is presented at chapter 6 so it is not repeated in this section.

7.8.4 Integration of components

The main interactions of VOLVO's use case components are depicted in Figure 42. The use case will be built as an IDS based use case.

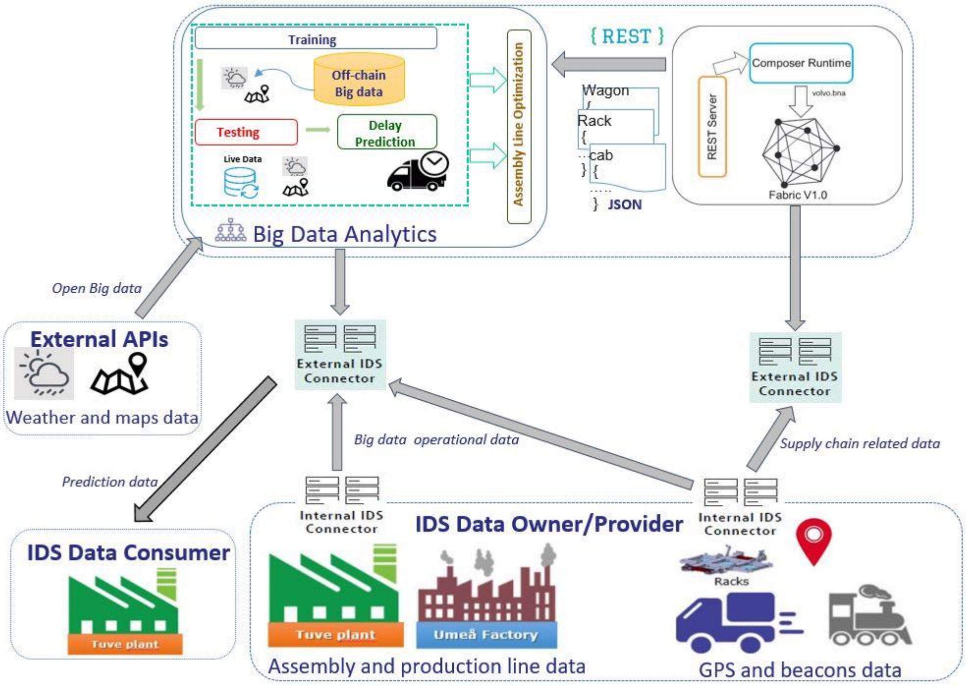

Figure 42 VOLVO Pilot High Level Architecture

The Big Data Analytic Tools and the Hyperledger Fabric are connected to data sources using IDS connectors. The data from assembly and production lines and the data from GPS receivers as well become available through internal IDS connectors. The tools are connected to these data using external IDS connectors.

Furthermore, the communication of Analytics Tools and External APIs is based on HTTP protocol and REST services. The External APIs are also providing their data in standard formats. Open APIs commonly use the JSON format.

The communication between the Big Data Analytics tool and the Hyperledger Fabric is based on well-known and standardised data formats and communication protocols. Their communication will be based on RESTful Web Services and HTTP protocol. The JSON format is used as the data format in this case. A JSON Schema has been defined. It is used as a common vocabulary that allows the developers of the two components to describe, to annotate and validate their JSON documents for the use case.

For the purposes of this use case and in alignment with the IDS architecture, the VOLVO will be the IDS Data Owner and Provider. Moreover, the Volvo's assembly line planning department will be the Data Consumer. Both Data Analytic tool and Fabric will be a kind of IDS App or Service.

7.9 Whirlpool whitegoods spare part sensing customer service factory 4.0

7.9.1 Short pilot overview

The main objective of the whirlpool trial is the optimisation of the spare parts planning and distribution process in the whole EMEA region through the full adoption of a prediction tool and technique that can help the organisation to better estimate when and where a spare part is needed.

The challenge is to create a forecasting tool that can be used to integrate different type of data into the same data model, to provide the data analysts with a good vision and understanding of what is happening on the different markets in terms of spare parts needs.

To achieve the objective of the pilot and the expected benefits for the company it has been necessary to create a centralised data model and database that incorporates, harmonises and aggregates all the relevant data extracted from different source systems and owned by different company departments.

The implementation of a big data analytic and prediction tool will be able to provide the company data analysts with the following set of outputs to react promptly to market needs (without waiting, like today, that the things happen):

1. Spare Parts Demand Forecast in Quantity and Frequency
2. Monitoring Tools & Demand Forecast Analysis
3. Quality Reports for Factory and Product Engineers, Market and Service Partners Training, Smart Appliance Predictive Maintenance Approach.

7.9.2 Short component description

A centralised Big Data Platform for Whirlpool has been designed for assuring internal data sharing in the pilot and data exchanges with external data consumers in the future.

The platform has the following components, as shown in Figure 43:

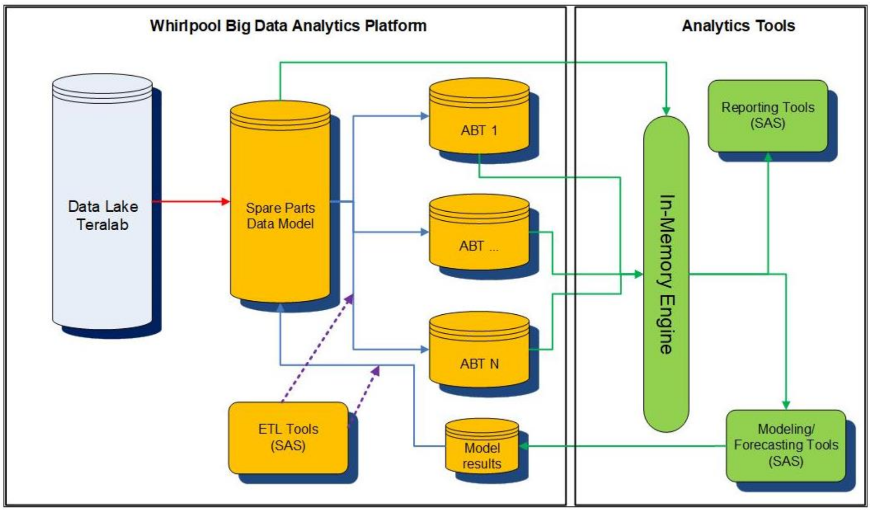

Figure 43 Whirlpool trial Big Data platform and analytics overview

The Data Lake is the input to the **Spare Parts Data Model**, which is the source for generating a variety of Analytical Base Tables used for studying, training and executing predictive models.

The analytical base tables are then loaded into SAS Viya **in-memory engine**, where **analytical tools** can perform data analysis of large amounts of data in few minutes. The in-memory engine is also used for interactive data exploration by end-user tools like SAS Visual Analytics.

7.9.2.1 Spare Part Data Model

A Spare Part Data Model is designed to manage data about any device that is composed by several spare parts; it enables generic forecasting and predictive analysis about spare parts and in particular:

- Spare parts sales forecasting. The forecast is based mainly on spare parts historical sales data, but they may also consider:
  - Device historical sales data
  - Service Order historical data
  - Factory functional and statistical tests
  - Environmental context (based on geolocalisation)
- Predictive maintenance on connected devices. These predictive models are used to detect potential anomalies on devices. They are based on physical parameters measured by sensors on connected devices.

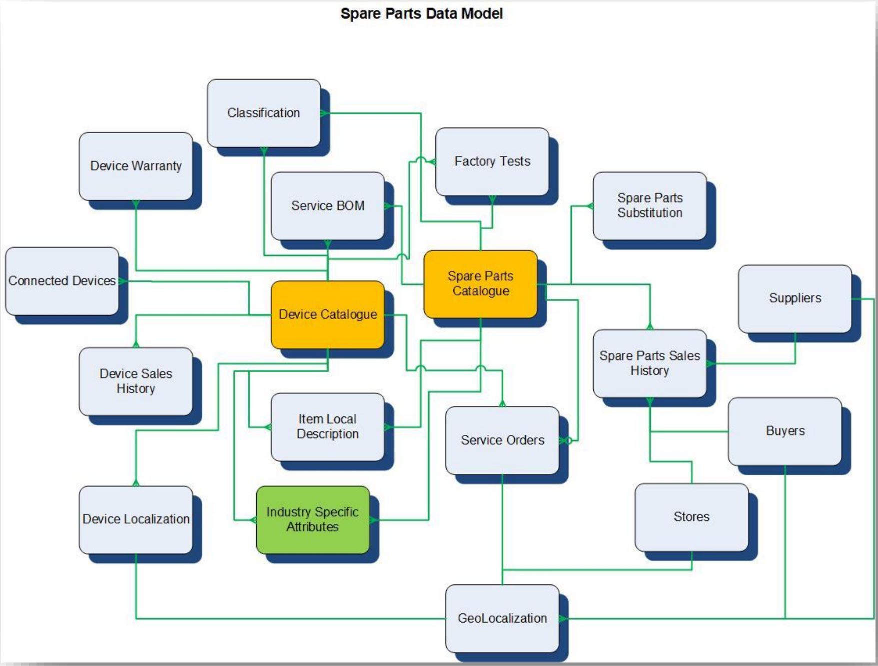

Figure 44 Spare part data model

7.9.2.2 Data Management tools

Data Management tools provided by SAS consist of:

- A set of database connectors to most of commercial database on the market, including Hadoop distributions (Cloudera, Hortonworks).
- SAS Data Management Languages (BASE SAS, SQL, FedSQL), which provides hundreds of functions for data processing and transformation
- Data Management Applications (Data Integration Studio, Enterprise Guide, and so on), which provides user interfaces for data processing flows generation, without code writing.

In Whirlpool Use Case, Data Management tools are used for:

- Extract, transform and load data from Whirlpool Data Lake into Spare Parts Data Model
- Extract, transform and load data from Spare Parts Data Model into Analytical Base Tables used by Analytical and Reporting Tools
- Load results from predictive models into Spare Parts Data Model

7.9.2.3 In-memory engine

SAS Analytical tools can analyse large amounts of data through an in-memory engine named CAS (Cloud Analytic Services); data must be loaded into CAS before analytical applications can work on them. Processes to load data into CAS are also developed with SAS Data Integration Studio.

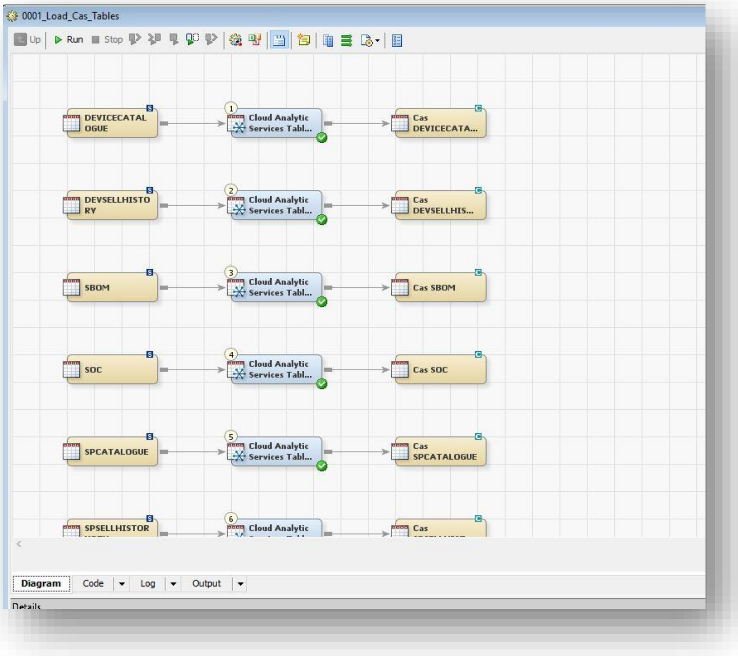

Figure 45 Loading data into CAS

For large scale experimentation, data will be provided on HDFS file system. All extraction and transformation processes will be executed by Hadoop (Hive or Impala) engine invoked by SAS programs. Data transfer between Hadoop and CAS will be performed through massive parallel process that will take advantage of SAS libraries on Hadoop and the co-located architecture. Analytical tools will make use of CAS in-memory engine.

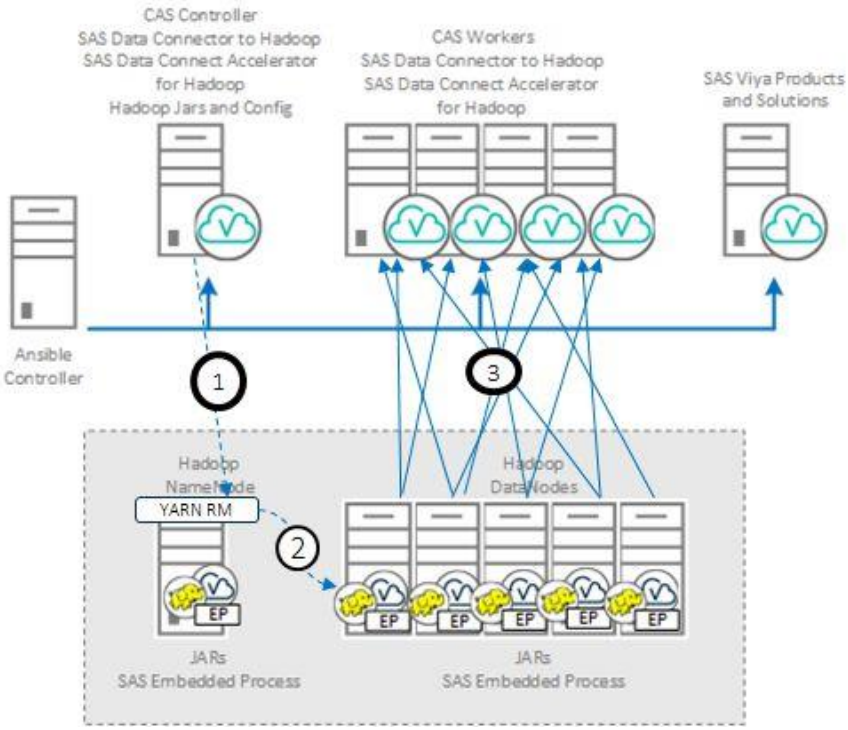

Figure 46 Overview of large-scale experimentation

7.9.2.4 Analytical tools

The Analytical tools for the Whirlpool pilot offer a set of procedures for the study, train and deploy of predictive models and forecasting applications:

- Visual Data Mining and Machine Learning
- Visual Forecasting

SAS Visual Data Mining and Machine Learning covers all aspects of machine learning and deep learning – from data access and data wrangling to sophisticated model building and deployment. In-memory, distributed processing handles large data and complex modelling, providing fast answers and efficient use of resources. Analytical capabilities include clustering, different flavours of regression, random forests, gradient boosting models, support vector machines, natural language processing, topic detection, these methods drive the identification of new patterns, trends and relationships between data attributes in structured and unstructured data, processing high velocity and high-volume data sets. Deep learning algorithms include deep neural networks, convolution neural networks for image classification and recurrent neural networks for improved text analysis.

SAS Visual Forecasting provides a resilient, distributed and optimised generic time series analysis scripting environment for cloud computing. This solution includes automatic forecast model generation, automatic variable and event selection, automatic parameter optimisation, automatic model selection and automatic forecast generation. It also provides advanced support for time series analysis (time domain and frequency domain), time series decomposition, time series modelling, signal analysis and anomaly detection (for IoT). The software determines the forecasting models that are most suitable for the historic data and that are more likely to predict the future.

Figure 47 shows an example of forecast model for spare part quantities:

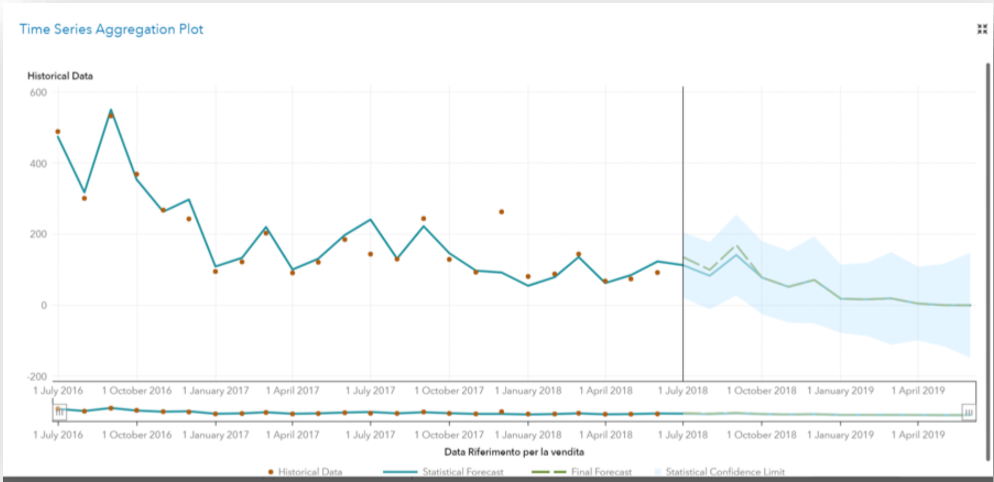

Figure 47 Example forecast model

7.9.2.5 Reporting Tools: Visual Analytics

SAS Visual Analytics provides a modern, integrated environment for governed discovery and exploration, with a large set of functionalities for examining trends and understanding relationships in data. Business users in Whirlpool can easily create and share reports and dashboards that monitor business performance and give insights from data. Users of all skill levels can visually explore data, use automated analysis and create visualisations while tapping into powerful in-memory technologies for faster computations and discoveries. This self-service solution scales to an enterprise wide level, putting data and analytics in the hands of more people. At the same time, governance capabilities help IT promote consistency and reuse.

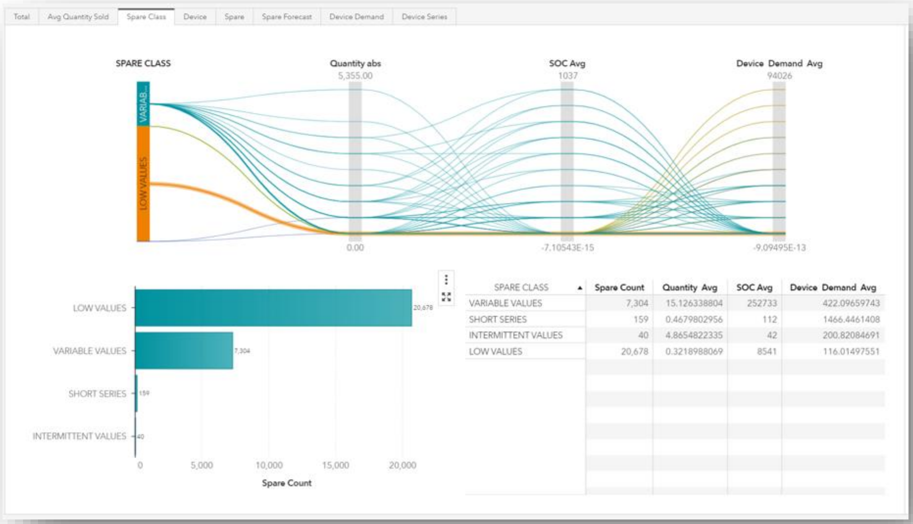

Figure 48 Visual Analytics

7.9.2.6 Vocabularies

Spare Parts Data Model contents will feed new ontologies and vocabularies that will be added to Boost Vocabulary provider: Protégé is a tool to create, modify and organise the conceptualisation of a domain of interest, namely an ontology. Ontologies can be described as formalised vocabularies of terms, shared by a community of users, which allow to define machine-interpretable basic concepts in a domain, together with the relations among them. For WHIRLPOOL pilot, Protégé is used to allow fast and easy access to information by external consumers who, in the future, could be authorised to access Whirlpool data.

7.9.3 Mapping to the RA

The components of the Whirlpool Big Data Platform described in the section 7.9.2 can be mapped to the BOOST 4.0 Reference Architecture; Figure 49 shows the architecture layers that are activated for Whirlpool pilot.

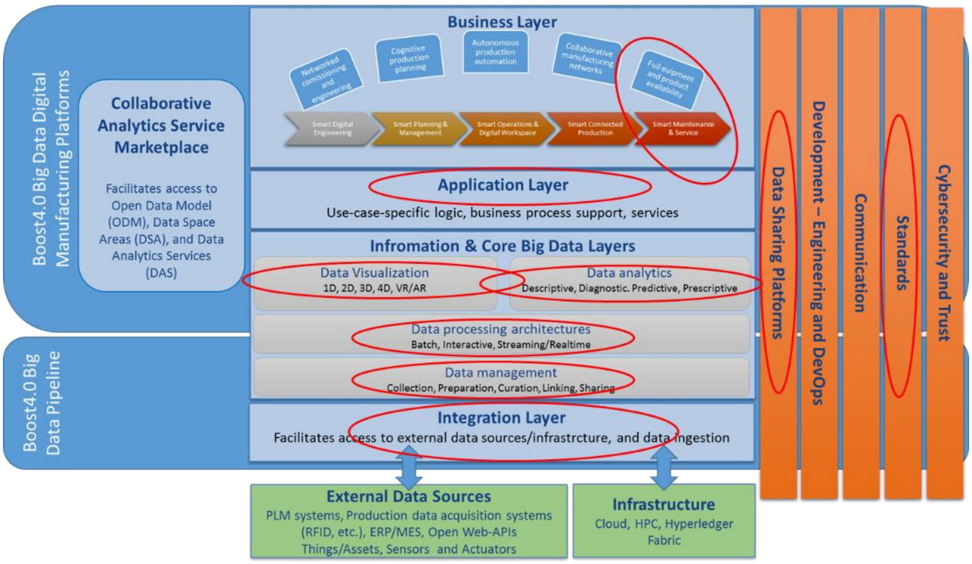

Figure 49 Relevant parts of the Boost4.0 RA for Whirlpool trial

In particular, SAS components overlap with the RA layers in the Whirlpool Big Data Platform in a unique correspondence; Figure 50 indicates the coverage of SAS software components of the architectural layers.

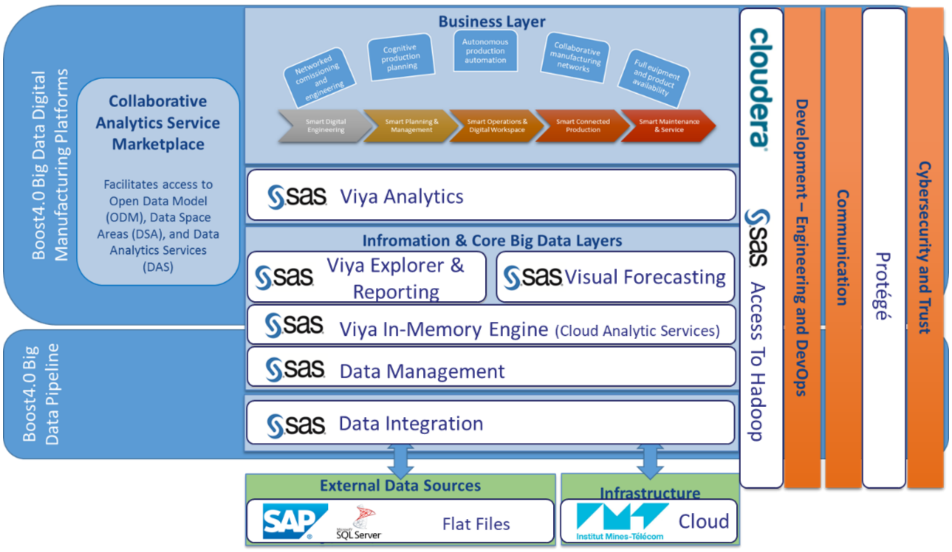

Figure 50 Overlap with SAS solutions

In particular, the following matching has been identified:

- The centralised Big Data Platform for Whirlpool covers the **Data Sharing Platform** layer
- The Spare Part Data Model designed for Whirlpool pilot overlaps with the **Data Management** layer

- The **Data Processing** layer is associated to ETL Tools for ABT creation and in-memory engine for fast data analysis on Big Data
- SAS Analytic Tools (Visual Data Mining and Machine Learning, Visual Forecasting) overlap with **Data Analytics Layer**, while SAS Reporting Tool (SAS Visual Analytics) covers the **Data Visualisation layer**
- The application of the forecasting analytic models to Whirlpool spare part data can be mapped to the **Application Layer**
- Protégé, the ontology editor, covers the **Standard** vertical component.

7.9.3.1 Data Ingestion

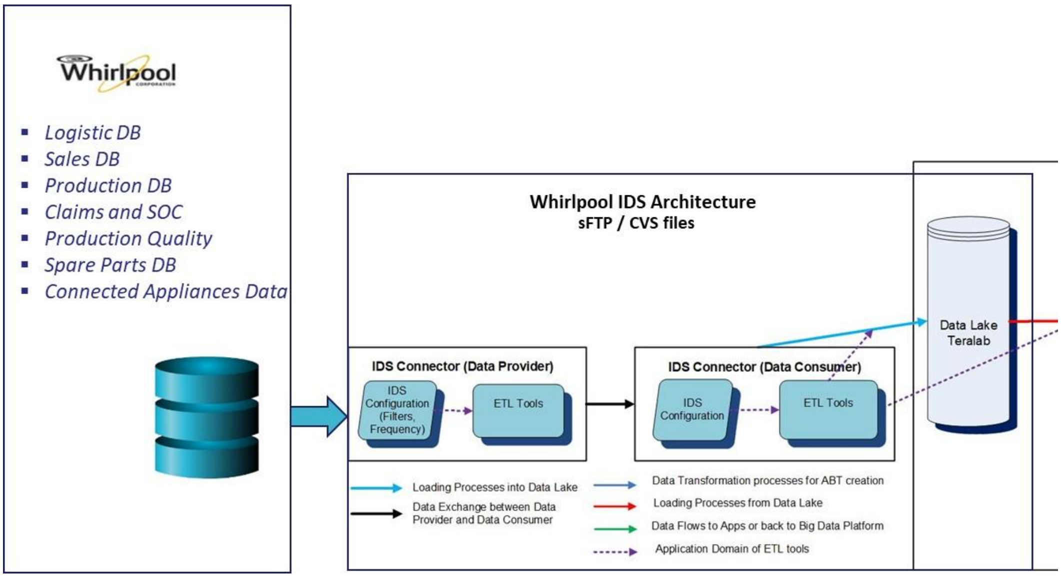

Figure 51 Data Ingestion

Spare parts data are collected by Whirlpool from data sources (legacy systems (SAP) and private clouds for connected devices) and loaded into a Data Lake in a Big Data Environment. Data are related to:

- Spare Parts Catalogue
- Spare Parts Sales History
- Device Catalogue
- Device Sales History
- Connected Devices Data
- Service Orders (with Spare Parts substitution)
- Service Bill of Material (association between Devices and their Spare Parts)
- Device Functional Tests.

Data provider components consists of set of extraction procedures from legacy systems, explicitly developed to get data relevant to the Spare Part Data Model.

Extraction procedure will be invoked by service calls from data consumer. Service calls may specify:

- Extraction procedure name
- Data Scope
- Filters
- Output format
- Output destination.

After a data extraction, launched after a service call, is completed, the data provider service will transfer data to the destination specified by the data consumer. At the end of the process, the data provider will notify data consumer of data availability or, in case of errors, of problems during data extraction/transfer.

7.9.4 Integration of components

7.9.4.1 Integration between Data Sources and Big Data Environment (Data Lake)

Integration between data extraction procedures from data sources and loading procedures into the Data Lake will be implemented through:

- REST services communication for exchanging request and messages
- Secure File transfer channels.

The following picture describes this process:

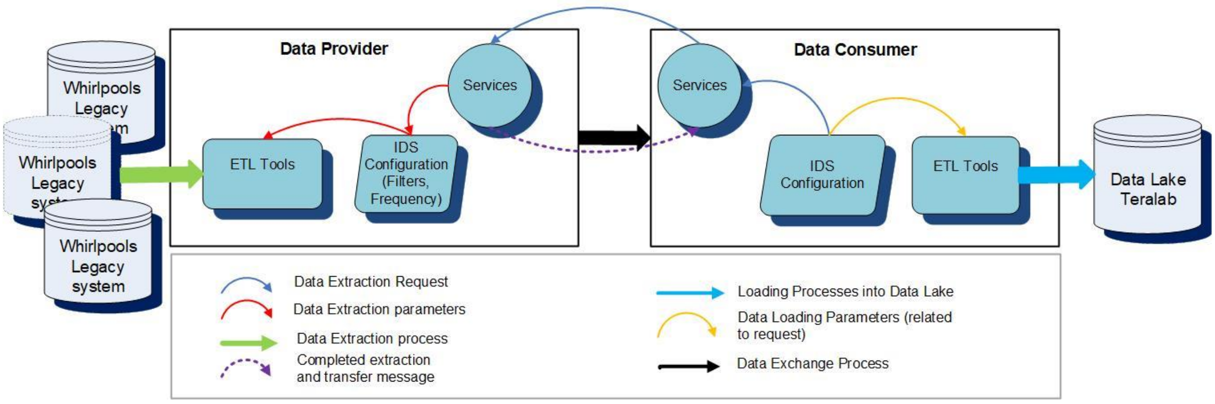

Figure 52 Integration between Data Sources and Big Data Environment

This sequence of steps describes the above process:

- Data Consumer asks data provider for data extraction. Data to be extracted and other extraction parameters are read from the IDS configuration and then passed to data provider as parameters of the service call

- Data provider saves request parameters into its own IDS configuration and starts ETL tools for data extraction
- As soon as data are available, data are transferred to data consumer. After transfer process is completed, data provider notifies data consumer about data transfer completion or potential errors.

After receiving success notification from data provider, data consumer starts ETL processes to load transferred data into the Data Lake. Parameters for ETL process are read from IDS configuration.

7.10 Benteler predictive factory 4.0

7.10.1 Short pilot overview

The pilot will implement a set of interconnected processes for Smart Maintenance and Services to boost productivity, reduce cost, and enhance quality. This includes:

- Condition monitoring of production machinery
- Prediction of plant failures based on data-driven models trained on historic big data and operate on real-time data streams.
- Fault diagnostics and prescriptive analytics based on data-driven models trained and operating on historic big data for assisting maintenance actions in case of plant failures by matching fault patterns and inferring proper maintenance actions.
- Concept and evaluation for integrated resource planning of equipment for maintenance process optimisation by interconnecting manufacturing, maintenance and diagnostic/predictive analytics data.

The project will consider Benteler's An der Talle factory as the focus of this pilot since An der Talle is Benteler's largest component production site with the highest turnover. At An der Talle, automotive chassis and structure components are produced by highly automated production lines. The factory's infrastructure entails equipment for joining, hot and cold forming, more than 500 robots, as well as special equipment for milling and laser cutting.

7.10.2 Short component description

7.10.2.1 Fault Detection Tool

Fraunhofer IEM will provide rule-based fault detection methods based on the knowledge from the previous analysis of processes and systems that will be augmented by online learning through data during runtime of the software. In the context of Benteler's use case, ATLANTIS will provide a fault detection tool (FDT). This tool will be supported by an outlier detection algorithm on streaming data, to analyse data from IoT devices and machine data and detect potential faults in the production line at real time.

For the FDT, a state-of-the-art algorithm for outlier detection on streaming data [44] will be used. This algorithm is able to analyse efficiently Big Data. The algorithm is provided in different programming languages like C++, Java and Scala. The latter is compatible with the Flink analytics [46] and Big Data software. The outlier detection algorithm will be able to utilise domain knowledge to pre-process/clear the incoming data and provide more accurate results. A rule-based mechanism provided by Fraunhofer IEM, will be also coupled with the outlier detection algorithm, as a backup mechanism, to further prevent unwanted events. The focus will be on solutions that are as unsupervised as possible from the machine learning perspective.

7.10.2.2Fault Prediction Tool

ATLANTIS will also provide a predictive maintenance tool (PMT) to predict failures in the production process, based on prior knowledge. For the PMT, planned to be applied an approach similar to the failure prediction algorithm using post flight reports [47], where a supervised machine learning algorithm is used (random forests), together with a risk function to form a regression problem. The algorithm analyses failure events from known failure modes. To provide a more flexible solution, in cases where the fault events are not available, technical providers are planning to utilise motif detection (Matrix Profile [48]) and/or outlier detection algorithms to produce such events that will be fed to an ensemble predictive algorithm. The solutions will be tailored to real-world Industry4.0 settings, where the fault events are very rare compared to the normal events, and there is no crisp knowledge as to which factors lead to/cause a problem.

Fraunhofer IEM will provide a semi-automated online-learning for predictive maintenance. The specified failures for both the business cases (oil leakage for the Hydraulic Press and metal jamming for the Scrap Belt) are rare events. Supervised learning methods are very difficult to train in the early phases, since no labels are available for training data sets. Rules can be deducted from the expert knowledge gathered in the analysis phase, to created unsupervised, rule-based fault detection methods. The output of these methods, combined with additional assessment by experts, can then be used for labelling of the time series data. The resulting labelled data can then be used to train supervised classifiers during runtime of the algorithm.

The proposed models need to be able to handle online streaming data to provide up-to-date input and early alerts to the maintenance staff. Batch processing solutions can also be applied in a periodic manner, as a secondary mechanism to strengthen the results dependability. The streaming processing of imbalanced data is an open research challenge. The proposed solution is important to be able to adapt to changes in the application environment and to be easily applied on new production sites. Unsupervised and Reinforcement learning provide self-training capabilities that address this demand.

7.10.2.3 DSS – Fusion Mechanism

Traditionally, the core of a Decision Support System (DSS) is a rule-based mechanism, which handles manually defined actions based on pre-specified static threshold violations. Such tools are used to assess the performance of the machines, to diagnose failures and overall to improve the maintainability and the operational efficiency of the production line.

ATLANTIS currently has an implementation of such a tool as a finite state machine, and there are plans extend the functionality within the scope of this project. It is proposed to extend the traditional notion of the DSS, by using data mining processes for Big Data analysis, IoT and machine learning techniques, to provide a DSS tool that is compatible with the key Industry4.0 element called smart maintenance. The goal is to create a tool that is able to diagnose, handle (to a certain amount, autonomously) and potentially predict future unwanted incidents (like machine failures.

The vision is to combine the output from maintenance experts, fault detection and prediction algorithms using an extendable fusion mechanism inside the DSS tool to provide easily interpretable and useful results. The solution will include combinations that consider the hybrid usage of different models. The outputs of multiple models can be fused by assigning a particular probability to each individual model or using a parametric meta-model. The evaluation of the predictive models needs to take into consideration the imbalanced domain.

7.10.3 Mapping to the RA

Figure 53, presents the mapping of the Benteler's use case to the Boost4.0 Reference Architecture. Starting from the bottom of the Figure, Benteler's engineers have already implemented the communication between the external data sources (for example machinery of a hot forming production line) and the infrastructure (PLCs, server deployment) of the Integration Layer.

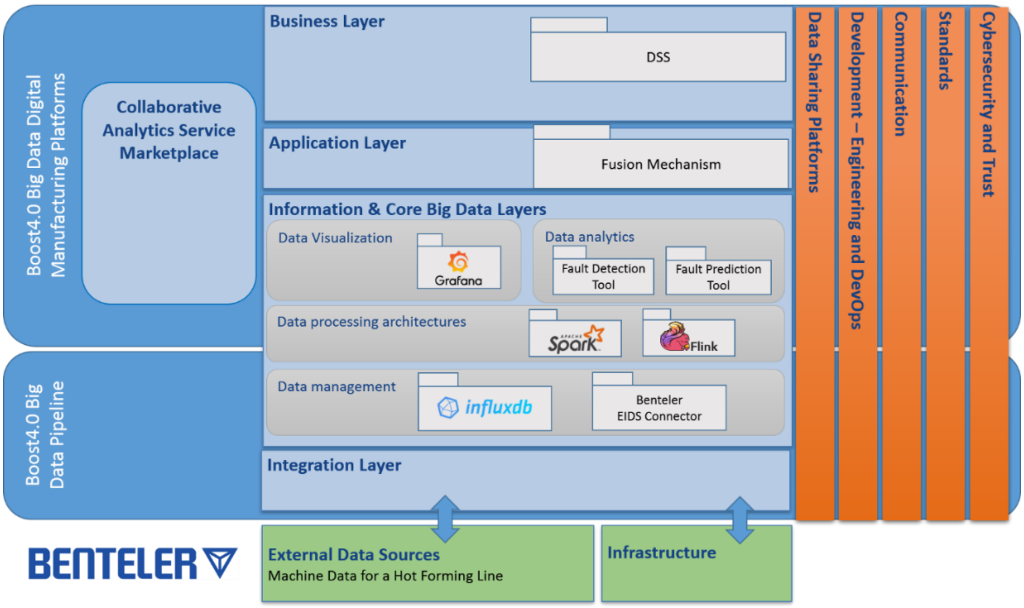

Figure 53 Mapping to the RA

Considering the Information & Core Big Data Layers, artefacts from the EIDS ecosystem will be the pillars for the implementation of the Industry 4.0 solutions that are proposed in this pilot. More specifically, in the Data management layer the machine data are stored in an InfluxDB timeseries database. Local InfluxDB deployments for each Benteler plant communicate with a Central Data Base (CDB) mirroring their data. The CDB is embedded into a sandboxed environment called Developers Space (DS), which is used for the implementation of Industry 4.0 solutions before their actual deployment in each of Benteler's plant. In the DS an EIDS connector will be implemented and deployed, which will communicate with EIDS AppStore for the fetching and the deployment of the applications located in the Data analytics layer (Fault Detection Tool (FDT) and Fault Prediction Tool (FPT)). The tools of the Data analytics layer will be able to process Big Data. They rely on technologies like Spark or Flink which can be made available in the Data processing architecture layer. The output of the tools will be directed to the CDB, which is connected with a Grafana visualisation platform (Data visualisation layer).

As mentioned in section 7.10.2, both the FDT and FPT will form a single solution through a fusion mechanism placed in the Application layer in Figure 53 Mapping to the RA. This mechanism will combine the output of the data analytics tools to provide a single output to the DSS, which is placed in the Business layer as it will process the obtained information from the data processing to extract business information like the integrated maintenance resource planning.

7.10.4 Integration of components

Component integration requires both the correct choice of a communication interface between components, as well as the correct specification of communication. There are proven protocols and standards which facilitate the definition of communication between components. One example is the paradigm of Representational State Transfer (REST) [88] for machine-to-machine communication. It is often applied to standard protocols like HTTP(S) which imply security layers (SSL) for component integration. Standard data formats used for integration include JSON or XML.

The components that will be developed within the BENTELER pilot factory, namely the Fault Detection Tool (FDT), Fault prediction tool (FPT) and Decision support system (DSS), also rely on standard methods, protocols and formats for integration with each other as well as with legacy components. The FDT and FPT at the analytics layer integrate with the InfluxDB at data management layer via ResultSet Objects. This is the InfluxDB response format transmitted via HTTP protocol. This includes data queries from the FDT and FPT to the InfluxDB, as well as writing back results into the InfluxDB, so that the legacy integration of Grafana and InfluxDB can be used for simple data visualisation. Components within the data processing architectures (for example SPARK, Flink) can be called optionally by the FDT and FPT algorithms.

The integration of the FDT and FPT with the DSS is not technically defined for the moment. It includes the transfer of multiple detection and prediction model outputs to the DSS, as well as a resulting time stamp of the fused fault detection / prediction result or a corresponding diagnosis and recommended action, which must be applied within the business process.

8 Conclusions

The document presents the work on the Boost 4.0 reference architecture which brings the big data dimension to the manufacturing domain. In order to facilitate a common understating within the project partners and support the pilots in identifying Big Data Assets required in order to fulfil their requirements a model with several layers and sub-layers, dimensions and cross-cutting aspects was developed. The model also brings together aspects from existing Reference Architectures and models targeting either Big Data (BDVRM, NBDRA) or the manufacturing domain (RAMI4.0, IIRA) and thus a clear alignment has been achieved.

Additionally, a mapping of significant Boost 4.0 project technologies/frameworks (BDE, IDS, Hyperledger Fabric) has been carried out in order to identify how and where these components fit. And how the pilots may leverage such components and integrate them into their solutions.

Finally, as last step the pilots were asked to validate the Boost 4.0 RA by placing their solutions within the framework of the Boost 4.0 RA and highlighting its capability of mixing together proprietary components and open-source components.

9 Appendix

9.1 Boost 4.0 components

This paragraph provides an overview of all the components used in the different pilot solutions and make-up the Boost 4.0 ecosystem. The components are organised per layer and come from all the different pilots.

9.1.1 Integration layer components

Table 5 Integration layer components

Component name	Description
IDS module provided by FIWARE	The FIWARE implementation of the IDS connector can be used in a variety of scenarios to facilitate data sharing with third parties.
VW Connector to EIDS	Allows the use of data streams from geographically distributed manufacturing processes in VW.
FILL machine abstraction layer	Integration of different data sources based on Kepware solution.

9.1.2 Information and Core Big Data layer components

Table 6 Information and Core Big Data layer layers components

Component name	Description	Sub-layer
Spark	Stream (real-time) processing of data.	Data processing architectures
Hadoop	Batch (historical) processing of data.	Data processing architectures
RabbitMQ	Streaming Text Oriented Messaging Protocol, Message Queuing Telemetry Transport, and other protocols	Data processing architectures
CERTH IoT Platform	Enables Big Data Storage, and supports real-time monitoring services • Supports multiple continuous connected IoT devices • Offers Real-time Predictive Analytics services	Data management

CERTH Data analytics	Support predictive analytics models and techniques that will enable the forecasting of cabs arrival time.	Data analytics
Benteler EIDS connector	Communicates with EIDS AppStore for the fetching and the deployment of the applications	Data management
InfluxDB	Open-source time series database	Data management
Flink	Open source stream processing framework	Data processing architectures
Grafana	Open-source, general purpose dashboard and graph composer, which runs as a web application	Data visualisation
Fault Detection Tool	Rule-based fault detection methods	Data analytics
Fault Prediction Tool	Predictive maintenance tool (PMT) to predict failures in the production process, based on prior knowledge	Data analytics
Kafka	An open-source stream-processing software platform	Data processing architectures
Predictive maintenance module	Uses machine learning to; for example, analyse the wear of ball screws as part of FILL milling machines.	Data analytics
Predictive Quality	Used for improvement on Fall Off Rate	Data analytics
Scilab	An open source software for numerical computation providing powerful computing environment for engineering and scientific applications.	Data analytics, Data visualisation
I2Tracking	Provides seamless integration between positioning technologies matching each event to a specific location and determining the different states of the objects that are being moved around.	Data analytics, Data visualisation
Scalable industrial data	Creates a modular pipeline of data from its acquisition to its external sharing	Data Management

Data Management Tools - SAS	Tool is used to define the ETL process to load data into the Spare Parts Data Model.	Data Management
Analytics tool – SAS	Effectively model and forecast time series in large scale.	Data analytics
Reporting Tools - SAS	Facilitates data exploration and result viewing	Data visualisation

9.1.3 Business and application layer components

Table 7 Business and application layer components

Component name	Description
Zero Defect Manufacturing (ZDM) powered by massive metrology	Provides the tools necessary to understand and analyse the position and activity of all factory assets in real time, as well as monitor all network activity.
DSS – Fusion Mechanism	A rule-based decision support tool extended to leverage data mining processes for Big Data analysis, IoT and machine learning techniques.
ESI Cloud platform	Virtual prototyping and data analytics for casting simulation
INENDI Inspector	INENDI Inspector solution is a software solution that allows any user to handle large amounts of structured or semi-structured data and perform a very rich and deep investigation of the data.
Digital twin plug-in	A plug-in of Visual Components allows a live visualisation of the current machine state (for example motion visualisation).
M3 Software	A high-performance software for capturing and analysing point clouds.
M3 Workspace	A cloud-based metrology software that synchronises with the main M3 Metrology software which allows for the automated uploading of any metrology results straight to the cloud
M3 analytics	A powerful tool that enables the visualisation, the statistics analysis and the reporting operations related to all the data stored in the cloud by means of several algorithms and computational components

MBDVIdia and MBDVidia	Tools enabling visualisation of the Product Manufacturing Information Data (PMI) with the semantic representation allowing the user to achieve high level of machine to machine communication reducing the human factor to the minimum.

9.1.4 Other layers or aspects

Table 8 Components belonging to other layer/aspects of the RA

Component name	Description	Layer/aspect
FILL Gate	Enables sharing of data between FILL customers and FILL central factory.	Data sharing platforms
redborder	A real-time NTA (Network Traffic Analysis) and active cybersecurity platform based on Big Data and Open Source	Cybersecurity and Trust

10 References

[1] Reference Architectures and Open Group Standards for the Internet of Things, http://www.opengroup.org/iot/wp-refarchs/p2.htm

[2] Michelle Knight, Data Curation 101: The What, Why, and How (2017), http://www.dataversity.net

[3] RAMI4.0: https://www.zvei.org/en/subjects/industry-4-0/the-reference-architectural-model-rami-40-and-the-industrie-40-component/

[4] Loukides, Mike (2012). "What is DevOps?", http://radar.oreilly.com/2012/06/what-is-devops.html

[5] Gartner Market Trends: DevOps – Not a Market, but Tool-Centric Philosophy That supports a Continuous Delivery Value Chain (Report). Gartner. 18 February 2015.

[6] Edwards, Damon. "Integrating DevOps tools into a Service Delivery Platform". dev2ops.org.

[7] Seroter, Richard. "Exploring the ENTIRE DevOps Toolchain for (Cloud) Teams". infoq.com.

[8] Theakanath, Thomas. "DevOps Stack on a Shoestring Budget". devops.com.

[9] "Stronger DevOps Culture with Puppet and Vagrant". Puppet Labs.

[10] "Gartner IT Glossary – Application Release Automation (ARA)". Gartner.

[11] Application Release Automation Market Forecast (Report). Forrester. 2015.

[12] http://www.bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf

[13] https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-6r1.pdf

[14] RAMI4.0: https://www.zvei.org/en/subjects/industry-4-0/the-reference-architectural-model-rami-40-and-the-industrie-40-component/

[15] IIRA: https://www.iiconsortium.org/IIRA.htm

[16] S. Auer, S. Scerri, A. Versteden, E. Pauwels, A. Charalambidis, S. Konstantopoulos, J. Lehmann, H. Jabeen, I. Ermilov, G. Sejdiu, A. Ikonomopoulos, S. Andronopoulos, M. Vlachogiannis, C. Pappas, A. Davettas, I. A. Klampanos, E. Grigoropoulos, V. Karkaletsis, V. de Boer, R. Siebes, M. N. Mami, S. Albani, M. Lazzarini, P. Nunes, E. Angiuli, N. Pittaras, G. Giannakopoulos, G. Argyriou, G. Stamoulis, G. Papadakis, M. Koubarakis, P. Karampiperis, A. N. Ngomo, and M. Vidal, "The BigDataEurope Platform – Supporting the Variety Dimension of Big Data," in 17th International Conference on Web Engineering (ICWE2017), 2017.

[17] J. Lehmann, G. Sejdiu, L. Bühmann, P. Westphal, C. Stadler, I. Ermilov, S. Bin, N. Chakraborty, M. Saleem, A. N. Ngonga, and H. Jabeen, "Distributed Semantic Analytics using the SANSA Stack,"; in _Proceedings of 16th International Semantic Web Conference - Resources Track (ISWC'2017), 2017.

a. Ermilov, J. Lehmann, G. Sejdiu, L. Bühmann, P. Westphal, C. Stadler, S. Bin, N. Chakraborty, H. Petzka, M. Saleem, A. N. Ngonga, and H. Jabeen, "The Tale of Sansa Spark," in Proceedings of 16th International Semantic Web Conference, Poster & Demos, 2017.

[18] G. Sejdiu, I. Ermilov, J. Lehmann, and M. Nadjib-Mami, "DistLODStats: Distributed Computation of RDF Dataset Statistics," in Proceedings of 17th International Semantic Web Conference, 2018.

[19] Hands-On Blockchain with Hyperledger, N. Gaur et al., 2018

[20] T. Swanson. Consensus-as-a-service: A brief report on the emergence of permissioned, distributed ledger systems. Report, available online, Apr. 2015. URL: http://www.ofnumbers.com/wp-content/uploads/2015/04/Permissioned-distributed-ledgers.pdf.

[21] Unified Threat Management". Gartner. Archived from the original on 13 Jul 2017. Retrieved 11 December 2017.

[22] Oppliger, Rolf (May 1997). "Internet Security: FIREWALLS and BEYOND". Communications of the ACM. 40 (5): 94.

[23] Gartner: Defining Intrusion Detection and Prevention Systems". Retrieved 2016-09-20.

[24] SIEM: A Market Snapshot". Dr.Dobb's Journal. 5 February 2007.

[25] Delfs, Hans & Knebl, Helmut (2007). "Symmetric-key encryption". Introduction to cryptography: principles and applications. Springer.

[26] P. Kinnucan (October 1978). "Data Encryption Gurus: Tuchman and Meyer". Cryptologia. 2 (4): 371.

[27] "Triple DES Encryption". IBM. Retrieved 2010-05-17.

[28] Daemen, Joan; Rijmen, Vincent (March 9, 2003). "AES Proposal".

[29] Lucian Constantin (14 May 2014). "Microsoft continues RC4 encryption phase-out plan with .NET security updates". ComputerWorld.

[30] Kent, S.; Atkinson, R. (November 1998). IP Encapsulating Security Payload (ESP). IETF. doi:10.17487/RFC2406. RFC 2406.

[31] Stallings, William (1999-01-01). Cryptography and Network Security: Principles and Practice. Prentice Hall. p. 164. ISBN 9780138690175.

[32] The Original RSA Patent as filed with the U.S. Patent Office by Rivest; Ronald L. (Belmont, MA), Shamir; Adi (Cambridge, MA), Adleman; Leonard M. (Arlington, MA), December 14, 1977, U.S. Patent 4,405,829 .

[33] Diffie, Whitfield; Hellman, Martin E. (November 1976). "New Directions in Cryptography".

[34] Craig Stuntz (2010-03-18). "What is Homomorphic Encryption, and Why Should I Care?".

[35] Berti, Hansche, Hare (2003). Official (ISC)² Guide to the CISSP Exam. Auerbach Publications. p. 379. ISBN 0-8493-1707-X.

[36] Eldon, Eric (2009-04-14). "Single sign-on service OpenID getting more usage". venturebeat.com. Retrieved 2009-04-25.

[37] "RFC 6749 - The OAuth 2.0 Authorisation Framework". Retrieved 2016-05-15.

[38] Forouzan, B.A. (2000). TCP/IP: Protocol Suite (1st ed.). New Delhi, India: Tata McGraw-Hill Publishing Company Limited.

[39] Mason, Andrew G. (2002). Cisco Secure Virtual Private Network. Cisco Press. p. 7.

[40] Rivest, R. (April 1992). "Step 4. Process Message in 16-Word Blocks". The MD5 Message-Digest Algorithm. IETF. p. 5. sec. 3.4.

[41] Florent Chabaud, Antoine Joux: Differential Collisions in SHA-0. CRYPTO 1998. pp56–71.

[42] Douglas R. Mauro & Kevin J. Schmidt. (2001). Essential SNMP (1st ed.). Sebastopol, CA: O'Reilly & Associates.

[43] RMON1: RFC 2819 - Remote Network Monitoring Management Information Base

[44] Hofstede, Rick; Čeleda, Pavel; Trammell, Brian; Drago, Idilio; Sadre, Ramin; Sperotto, Anna; Pras, Aiko. "Flow Monitoring Explained: From Packet Capture to Data Analysis with NetFlow and IPFIX". IEEE Communications Surveys & Tutorials. IEEE Communications Society. 16 (4): 28.

[45] D. Georgiadis, M. Kontaki, A. Gounaris, A. N. Papadopoulos, K. Tsichlas, Y. Manolopoulos. "Continuous outlier detection in data streams: an extensible framework and state-of-the- art algorithms." In SIGMOD International Conference on Management of Data, pp. 1061-1064. ACM, 2013.

[46] P. Carbone, A. Katsifodimos, S. Ewen, V. Markl, S. Haridi, K. Tzoumas. "Apache flink: Stream and batch processing in a single engine." Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 36, no. 4 (2015).

[47] P. Korvesis, S. Besseau, M. Vazirgiannis. "Predictive Maintenance in Aviation: Failure Prediction from Post Flight Reports." In Data Engineering (ICDE), pp. 1414-1422, 2018

[48] Yeh, Chin-Chia Michael, Nickolas Kavantzas, and Eamonn Keogh. "Matrix profile IV: using weakly labeled time series to predict outcomes." Proceedings of the VLDB Endowment 10, no. 12 (2017): 1802-1812.

[49] DCC Curation Lifecycle Model http://www.dcc.ac.uk/resources/curation-lifecycle-model.

[50] Ipsita Pattnaik, "Advantages of Using Hadoop", https://blog.e-zest.com/advantages-of-using-hadoop (2016)

[51] DataFlair Team, "13 Big Limitations of Hadoop & Solution To Hadoop Drawbacks" (2017)

[52] Jim Scott, "Stream Processing Everywhere – What to Use?"

[53] Unified eXchange Platform, https://cyber.ee/products/secure-data-exchange/materials/uxp-technical-whitepaper-2018.pdf

[54] International Data Spaces, https://www.internationaldataspaces.org/

[55] Open Micka, http://micka.bnhelp.cz/

[56] European Open Science Cloud (EOSC), https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

[57] Microsoft Team Foundation Server is an enterprise-grade server for teams to share code, track work, and ship software — for any language. https://visualstudio.microsoft.com/tfs/

[58] Git is an easy to learn and has a tiny footprint with lightning fast performance. It outclasses SCM tools like Subversion, CVS, Perforce, and ClearCase with features like cheap local branching, convenient staging areas, and multiple workflows. https://git-scm.com/

[59] AWS CodeCommit is a fully-managed source control service that makes it easy for companies to host secure and highly scalable private Git repositories. CodeCommit can be used to securely store anything from source code to binaries, and it works seamlessly with existing Git tools. https://aws.amazon.com/codecommit/

[60] Subversion is an open source version control system. Founded in 2000 by CollabNet, Inc., the Subversion project and software have seen incredible success over the past decade. Subversion has enjoyed and continues to enjoy widespread adoption in both the open source arena and the corporate world. https://subversion.apache.org/

[61] Helix Core is an enterprise-class version control for faster, more collaborative development for teams of any size. With Helix Core, developers can use their preferred tools — like Git and Visual Studio — while enjoying lightning-fast workflows, support for all file types, and faster builds from a single source of truth. https://www.perforce.com/products/helix-core

[62] IBM Rational ClearCase is a software configuration management solution that provides version control, workspace management, parallel development support, and build auditing. https://www.ibm.com/us-en/marketplace/rational-clearcase

[63] Jenkins is an open-source CI tool written in Java. Jenkins is a cross-platform CI tool and offers configuration both through GUI interface and console commands. Jenkins is extensible through plugins. Besides extensibility, Jenkins prides itself on distributing builds and test loads on multiple machines. It is published under MIT license so it is free to use and distribute. https://jenkins.io/

[64] TeamCity offers all the features in its free version, but it is limited to the 100 build configurations and 3 build agents. Additional build agents and build

configurations need to be purchased. Out of the box, TeamCity works on many different platforms and has the support for wide variety of tools and frameworks. There are many publicly available plugins, developed both by JetBrains and third parties. Despite being the Java-based solution, TeamCity offers.NET support. https://www.jetbrains.com/teamcity/

[65] Travis CI, although it is mostly known its hosted solution, an on premise version in a form of enterprise package also exists. Travis CI is free for all open source projects hosted on the GitHub or for the first 100 builds otherwise. It supports a variety of different languages and a good documentation to back them up. https://travis-ci.org/

[66] Go Continuous Delivery, excluding the commercial support that ThoughtWorks offers, Go is free of charge. It is available for Windows, Mac, and various Linux distributions Go supports the concept of pipelines which makes the modelling of the complex build workflows easy. https://www.gocd.org/

[67] GitLab CI is hosted on GitLab.com, a free hosted service and it provides detailed git repository management with features like access control, issue tracking, code review, and so on GitLab CI integrates seamlessly with GitLab and it can easily hook projects using the GitLab API. Gitlab CI comes with both the open-source GitLab Community Edition and with the GitLab Enterprise Edition. https://about.gitlab.com/product/continuous-integration/

[68] Selenium with Robot Framework is an open source framework for automated testing. It supports multiple operating systems (Windows, Mac, and Linux) and browsers (Chrome, Firefox, IE, and Headless browsers). Test Scripts can be developed in various languages such as Java, Groovy, Python, C#, PHP, Ruby, and Perl. To enable continuous testing frameworks on top of Selenium are required. There are a number of tools and frameworks built on Selenium. With Robot, variety of automated tests can be run; both UI element based and API tests, with the assistance of Selenium Webdriver. http://robotframework.org

[69] Testsigma is a unified AI-driven test automation platform that integrates Continuous Testing in an end-to-end continuous delivery ecosystem. Testsigma provides agility to both functional and automation teams by eliminating the technical complexity of test automation. Testsigma allows for multiple open source and third-party integrations and supports thousands of test environments with different device/browser/OS combinations on the cloud to match with the dynamic testing requirements continuously. https://testsigma.com/

[70] IBM Rational Functional Tester is capable of Functional, API, Performance Testing and Regression testing. RFT creates automated functional tests by using record and playback and compares the actual result produced by the system at the time

of execution, with the expected result stored at the time of recording. IBM supports a wide variety of applications and offers integrations with multiple other tools in addition to its Rational Performance Tester, IBM Rational Quality Manager, and so on. https://www.ibm.com/us-en/marketplace/rational-functional-tester

[71] Tricentis Tosca applies a model-based testing approach and makes script maintenance easy. Tosca can run numerous tests a day continuously. Tricentis Tosca uses risk-based test design to suggest the most effective test cases and identify the risk contribution of each test cases. It maximises the reusability and increases maintainability with model-based test automation. https://www.tricentis.com/software-testing-tools/

[72] Unified Functional Test (UFT) is a cross-platform automation testing tool. It allows the execution of tests using both keywords and manual scripting. It supports integration with CI tools, provides web, mobile, and API testing, and has a smart object recognition function. UFT uses VBScript as the scripting language and requires knowledge in this domain. It has Reusable Test Components, which reduces redundancy in Test Steps, easy conversion of Manual Tests to Automated ones, shift-left compliance, and Test Execution on multiple machines and devices. https://www.microfocus.com/en-us/products/unified-functional-automated-testing/overview

[73] JFrog Artifactory is an artefact repository manager, which is entirely technology agnostic and fully supports software created in any language or using any tool. It is also the only enterprise-ready repository manager available that supports secure, clustered, high availability Docker registries. https://www.jfrog.com/confluence/display/RTF/Welcome+to+Artifactory

[74] Sonatype Nexus is a repository manager that allows for proxying, collecting, and managing dependencies. It makes it easy to distribute software. Internally, it allows developers to configure their build to publish artefacts to Nexus and they then become available to other developers. https://www.sonatype.com/nexus-repository-sonatype

[75] Apache Archiva is an extensible repository management software that helps taking care of personal or enterprise-wide build artefact repository. It supports build tools such as Maven, Continuum, and ANT. Archiva offers several capabilities, amongst which remote repository proxying, security access management, build artefact storage, delivery, browsing, indexing and usage reporting. https://archiva.apache.org/index.cgi

[76] XL Deploy is an application release automation tool that can deploy applications to various environments, all while managing configuration values that are specific to each environment. It is designed to make the process of deploying applications

faster, easier, and more reliable. https://xebialabs.com/integrations/xl-release-xl-deploy/

[77] Visual Studio Release Management enables the automation of complex deployments to a variety of target environments. https://docs.microsoft.com/en-us/azure/devops/pipelines/release/what-is-release-management?view=vsts

[78] IBM UrbanCode Release Automation manages the release of complex interdependent applications, infrastructure changes and simultaneous deployments of multiple applications. This software enables planning, execution and tracking of a release through every stage of the lifecycle model. https://www.ibm.com/support/knowledgecenter/en/SS4GCC_6.2.1/com.ibm.urelease.doc/topics/intro_overview.html

[79] CA Release Automation is an enterprise-class, continuous delivery solution that automates complex, multi-tier release deployments through orchestration and promotion of applications from development through production. https://xebialabs.com/technology/automicreleaseautomation/

[80] ElectricFlow is an Adaptive Release Orchestration platform that enables teams to implement fast and adapt easily for unprecedented insight and control of all types of releases – at any scale. https://electric-cloud.com/products/electricflow/

[81] Ansible is an open source tool for automating configuration management, deployment, and orchestration. Ansible can automate the deployment and configuration of ephemeral instances, and remove them when no longer needed. It can perform auto-scaling for modern, scale-out applications. https://www.ansible.com/

[82] Chef is an automation platform that transforms infrastructure into code. Chef automates how infrastructure is configured, deployed, and managed across network. https://www.chef.io/chef/

[83] Puppet is an open-source software configuration management tool. It runs on many Unix-like systems as well as on Microsoft Windows, and includes its own declarative language to describe system configuration. https://puppet.com/products/open-source-projects

[84] SaltStack is a Python-based open-source configuration management software and remote execution engine. Supporting the "Infrastructure as Code" approach to deployment and cloud management. https://www.saltstack.com/

[85] New Relic is a software analytics product for application performance monitoring (APM) delivers real-time and trending data about web application performance. With end to end transaction tracing and a variety of color-coded charts and reports, APM visualises data, down to the deepest code levels. https://newrelic.com/

[86] AppDynamics provides a significant amount of detail to solve performance problems by using APM Tools and an analytics-driven approach. Its Application Performance Management solution baselines, monitors and reports on the performance of all transactions that flow through an app. https://www.appdynamics.com/
[87] Compuware APM combines Strobe, Abend-AID and ThruPut Manager. These can be used to identify inefficiencies as well as reduce hardware and software costs. https://compuware.com/application-performance-management/
[88] https://www.w3.org/TR/2004/NOTE-ws-arch-20040211/#relwwwrest
[89] Extract, transform, load (ETL), is the process used to pull data out of one source and put it into another data storage, the data is transformed to match the format of the second storage before sending it the data. On the other hand Extract, Load, Transform (ELT) stores the data directly into the second data storage and the transformation is handled by the second storage system.
[90] Apache Haddoop, https://hadoop.apache.org/
[91] Apache Storm, http://storm.apache.org/
[92] Apache Samza, http://samza.apache.org/
[93] Apache Spark, https://spark.apache.org/
[94] ApacheFlink, https://flink.apache.org/
[95] Docker, https://www.docker.com/
[96] BDE components, https://github.com/big-data-europe/README/wiki/Components
[97] https://hub.docker.com/r/bde2020/
[98] https://github.com/big-data-europe/mu-init-daemon-service
[99] https://github.com/big-data-europe/app-stack-builder
[100] https://github.com/big-data-europe/WorkFlow-Builder
[101] https://github.com/big-data-europe/WorkFlow-Monitor
[102] https://github.com/big-data-europe/app-bde-pipeline
[103] https://github.com/big-data-europe/app-swarm-ui
[104] HyperLedger Fabric, https://www.hyperledger.org/projects/fabric
[105] Apache NiFi, https://nifi.apache.org
[106] Apache HBase, https://hbase.apache.org
[107] Apache Phoenix, https://phoenix.apache.org
[108] OPC-UA, https://opcfoundation.org/about/opc-technologies/opc-ua
[109] RabbitMq, https://www.rabbitmq.com/
[110] https://wso2.com/
[111] https://github.com/IndustrialDataSpace/InformationModel
[112] R. Hecht – Nielsen, III.3 Theory of back propagation neural network, HNC, Inc. and University of California, San Diego, 1992, p. 65-93.

[113] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers. In: 5th Annual Workshop on Computational Learning Theory, ACM, (1992) 144–152.

[114] S.J. Lee, K. Siau, A review of data mining techniques. Ind. Manage. Data Systems: 101(1), (2001) 41–46.

[115] L. Breiman, Random forests – random features. Technical Report 567, Statistics Department, University of California, Berkeley, ftp://ftp.stat.berkeley.edu/pub/users/ breiman.

[116] B. Hajek, Random processes for engineers, Cambridge University Press, ISBN 978-1-316-24124-0, 2015.

[117] L.E. Buam, T. Petrie, Statistical Inference for probabilistic functions of finite state Markov Chains, The Annals of Mathematical Statistics, 37(6) 1966, 1554-1563

[118] Industry Ontology Foundry: https://sites.google.com/view/industrialontologies/home

[119] See Eric Porter's 1985 book "Competitive Advantage: Creating and Sustaining Superior Performance", published by Simon and Schuster, New York.