D1.2 – Benchmarking of Digital Technologies with potential to i4Q

WP1 – NEED: Industrial Scenarios and Requirements Analysis

i4Q has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958205.

Document Information

GRANT AGREEMENT NUMBER	958205	ACRONYM		i4Q
FULL TITLE	Industrial Data Services for Quality Control in Smart Manufacturing			
START DATE	01-01-2021	DURATION		36 months
PROJECT URL	https://www.i4q-project.eu/			
DELIVERABLE	D1.2 – Benchmarking of Digital Technologies with potential to i4Q			
WORK PACKAGE	WP1 – NEED: Industrial Scenarios and Requirements Analysis			
DATE OF DELIVERY	CONTRACTUAL	April 2021	ACTUAL	April 2021
NATURE	Report	DISSEMINATION LEVEL		Public
LEAD BENEFICIARY	CERTH			
RESPONSIBLE AUTHOR	Anastasios Karakostas (CERTH), Georgia Apostolou (CERTH)			
CONTRIBUTIONS FROM	1-CERTH, 3-IBM, 4-ITI, 5-KBZ, 7-IKER, 8-BIBA, 9-UPV, 10-TUB, 11-UNI			
TARGET AUDIENCE	1) i4Q Project beneficiaries; 2) industrial community; 3) other H2020 funded projects; 4) scientific community			
DELIVERABLE CONTEXT/ DEPENDENCIES	This document has no preceding documents or further iterations. However, the results of this document will be input to the RTD WPs tasks.			
EXTERNAL ANNEXES/ SUPPORTING DOCUMENTS	None			
READING NOTES	None			
ABSTRACT	Deliverable D1.2 aims to present and systematize knowledge about digital technologies, especially related to the manufacturing quality and data reliability, gathered from the i4Q project's partners and characterise the most recent developments. A multi-dimensional benchmarking instrument will be developed in order to support the i4Q design and development. Results of this analysis will be used to establish benchmarks for each of the dimensions presented in this document. To conclude the state-of-the-art of emerging and promising digital technologies will be also characterised.			

Document History

VERSION	ISSUE DATE	STAGE	DESCRIPTION	CONTRIBUTOR
0.1	28-Jan-2021	Draft	Table of contents and sections allocation to partners	CERTH, UPV
0.2	17-Feb-2021	Draft	Benchmarking Evaluation Framework defined	UPV, KBZ
0.3	21-Feb-2021	Draft	Digital Twins	UPV
0.4	26-Feb-2021	Draft	Fog/Edge Computing; IIoT	UPV
0.5	3-Mar-2021	Draft	Blockchain/ Hyperledger; Data Analytics	CERTH
0.6	10-Mar-2021	Draft	Big Data; Machine Learning	CERTH
0.7	17-Mar-2021	Draft	Security Technologies	IKERLAN, UPV
0.8	26-Mar-2021	Draft	Benchmarking Analysis among the Digital Technologies Application of the Digital Technologies in the i4Q Solutions	CERTH, IBM, IKERLAN, ITI, UNI, TUB, UPV, BIBA
0.9	31-Mar- 2021	Draft	Conclusions	CERTH
1.0	31-Mar-2021	Draft	Conclusions All sections final draft Deliverable ready for internal review	ALL
2.0	13-Apr-2021	Draft reviewed	Deliverable reviewed	IKER, DIN
3.0	26-Apr-2021	Draft	Final version sent to CERTH for the final quality check	ALL
4.0	27-Apr-2021	Draft	Addressing some minor comments	CERTH
5.0	28-Apr-2021	Final doc	Final quality check and issue of final document	CERTH

Disclaimer

Copyright message

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS/ACRONYMS

ABE	Attribute-Based Encryption
ACME	Automated Certificate Manamgement Environment
AI	Artificial Intelligence
API	Application Programming Interface
APPs	Applications
AR	Augmented Reality
AWS	Amazon Web Services
BFT	Byzantine Fault Tolerance
BIM	Building Information Modeling
BSD	Berkeley Software Distribution
CA	Certificate Authority
CAE	Computer-Aided Engineering
CMP	Certificate Management Protocol
CNC	Computer Numerical Control
CPS	Cyber-Physical System
CPU	Central Processing Unit
CRISP	Cryptographic Reduced Instruction Set Processor
DB	Data Base
DFM	Design for Manufacturability
DLT	Distributed Ledger Technology
DSA	Digital Signature Algorithm
DSS	Digital Signature Standard
DTC	Design to Cost
E2EE	End-to-end encryption
EDQC	Evolutional Data Quality Concept
EMV	Europay, Mastercard, and Visa
ERC	European Research Council
EST	Enrollment over Secured Transport
ETL	Extract, Transform and Load
EU	European Union

EVM	Ethereum Virtual Machine
FMI	Functional Mock-up Interface
FMU	Functional Mock-up Unit
GA	Grant Agreement
GE	General Electric
GIS	Geographic Information System
GPG	GNU Privacy Guard
GPIO	General Purpose Input/Output
HMI	Human-Machine Interface
HSM	Hardware Security Module
HTTP	Hypertext Transfer Protocol
IaaS	Infrastructure as a Service
IACS	Industrial and Automation Control System
IBE	Identity-Based Encryption
IBM	International Business Machine
IC	Integrated Circuit
ICT	Information and Communication Technologies
IDE	Integrated Development Environment
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
IIoT	Industrial Internet of Things
IoT	Internet of Things
IPCs	Inter-Process Communication
IPSec	Internet Protocol security
ISO	International Organization for Standardization
IT	Information Technology
JMV	Java Virtual Machine
JSON-RPC	JavaScript Object Notation - Remote Procedure Call
M2M	Machine-to-Machine
MCU	Microcontroller Unit
MEC	Multi-Access Edge Computing
MQTT	Message Queuing Telemetry Transport

NA	Not Applicable
NB-IoT	Narrow Band IoT
Nd	nDimentional
NFV	Network Functions Virtualization
NPL	Natural Language Processing
OEM	Original Equipment Manufacturer
ONOS	Open Network Operating System
OPC	OLE for Process Control
OS	Operating System
OSGi	Open Services Gateway initiative
OT	Operational Technology
PaaS	Platform as a Service
PCKS	Public-Key Cryptography Standards
PGP	Pretty Good Privacy
PKC	Public key cryptography
PKI	Public Key Infrastructure
PLC	Programmable Logic Controller
PLM	Product Lifecycle Management
R&D	Research and Development
RAN	Radio Access Network
REST	Representational state transfer
RFC	Request for Comments
RISC	Reduced Instruction Set Computing
ROM	Read Only Memory
RPC	Remote Procedure Call
RSA	Rivest, Shamir and Adleman
S/MIME	Secure / Multipurpose Internet Mail Extension
SAP	Systems, Applications, Products in Data Processing
SCADA	Supervisory Control And Data Acquisition
SCEP	Simple Certificate Enrollment Protocol
SDK	Software Development Kit
SDN	Software Defined Networking

SHM	Structural Health Monitoring
SILC	Secure Internet Live Conferencing
SME	Small Medium Enterprise
SoC	System on Chip
SOC	Security Operations Center
SOLMA	Scalable Online Machine Learning and Data Mining Algorithms
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
SSH	Secure SHell
SW	Software
TLS	Transport Layer Security
TMP	Trusted Platform Module
TPM	Trusted Platform Module
TRNG	True Random Number Generator
TSN	Time-Sensitive Networking
TXT	Text
VoIP	Voice over Internet Protocol
WISE	Wireless Sensor Networks
WP	Work Package
WPAN	Wireless Personal Area Network
WWW	World Wide Web
ZDMP	Zero Defects Manufacturing Platform
ZRTP	Z and Real-time Transport Protocol

Executive summary

The digital technologies landscape is increasingly dynamic along several dimensions, such as the technological infrastructure and architecture, business model, data/information management, and social interaction. It is then necessary a multi-dimensional benchmarking instrument to be developed, supporting the i4Q design and development. Task T1.2 systematizes knowledge about digital technologies (in general and, specifically, related to manufacturing quality and data reliability) gathered from the project partners that are specialists in the area and characterises the most recent developments. The information sources are academic and professional literature and all the results from European Union (EU) projects dedicated to this subject.

In deliverable D1.2, the state-of-the-art of emerging and promising digital technologies (e.g., Blockchain, hyperledger, fog/edge computing, data analytics, big data, machine learning, Industrial Internet of Things - IIoT, digital twin, etc.) is characterised. Each technology is described and analysed, according to its state, maturity, tools, EU project solutions, benchmarking and assessment, application of i4Q solutions. At the same time, an evaluation framework is defined for the comparative analysis of technologies' features. Six dimensions are used, namely; general, technological, business model (financials, regulatory, etc.), informational and social and various evaluation criteria, in order to characterize these technologies.

Results show that relevance and capability are the best-fulfilled criteria from the general criteria, while integration is the best fulfilled from the technological ones. Maturity and support are the best fulfilled from the business model criteria, need for data traceability among the informational criteria, and finally, social preferences and development-friendliness are the best-fulfilled among the social criteria. Relevance and capability are fulfilled by all technologies analysed at the highest possible level. Furthermore, Data Analytics, Machine Learning and Big Data are the three technologies that meet most criteria, with a quite good rating. Results are used to identify the best cases – the cases in which the technologies and tools meet the most criteria, as well as which of the solutions could be used by most tools – establishing benchmarks for each of the dimensions.

Moreover, when performing the matching between the technologies and their associated tools with the i4Q's solutions, the results are aligned with the results of the first analysis (matching of technologies and criteria). Data Analytics, Machine Learning, and Big Data are the technologies that meet most criteria with a quite good rating and are also the ones in which a greater number of tools will be used for the development of the i4Q solutions. Python libraries seem to be an appropriate data analytics/visualisation related tool for the development of 8 of the 19 i4Q solutions (in total there are 22 solutions but 5 of them are guidelines). In the case of the digital technology of machine learning, Python is the most selected tool to develop different i4Q solutions. Finally, the most important big data related tool selected for the development of i4Q solutions is Tensorflow, as four of the six i4Q solutions that will use this technology has chosen this tool.

Document structure

Section 1: The deliverable begins with the introduction in Section 1, where the context of the document is described, as well as the methods that will be used in order to collect the appropriate information for the analysis of the technologies and the development of i4Q solutions.

Section 2: In Section 2 the benchmarking evaluation framework is presented, as well as six proposed dimensions, and various criteria for the assessment of the technologies that will be analyzed and selected as relevant to fulfill the development objectives of the i4Q solutions.

Section 3: In Section 3 the digital technologies for manufacturing quality and data reliability are extensively addressed. Eight technologies are presented, including professional tools, benchmarking and assessment, as well as applications for i4Q solutions.

Section 4: Section 4 includes the benchmarking analysis among the digital technologies that were analyzed in Section 3.

Section 5: Section 5 focuses on the application of the digital technologies in the 22 proposed i4Q solutions.

Section 6: In this Section the conclusions of the deliverable are presented.

1. Introduction

The development of the deliverable D1.2 and the information it contains is based on the knowledge and background provided by all the beneficiaries of the project, not only of those who participate directly in WP1 (Work Package 1) and specifically in task T1.2, but also of those beneficiaries who will work in the BUILD Workpackages (WPs) since these WPs require a selection of the most appropriate technologies for the achievement of the objectives of the i4Q developed solutions. The beneficiaries' background involves:

- Its knowledge in the different technologies
- Its expertise and ownership of different solutions
- Its participation in different research and development projects that have used and/or develop any technologies that could be of interest to i4Q.

To collect this piece of information, besides periodical meetings, a collaborative process was defined and shared means were out at disposal of the project consortium (Annex I). Moreover, it is worth mentioning that the process of collecting information through shared documents is aligned with the process of collecting information of task T1.1.

In this case, in task T1.2, the different tools (Python libraries, R libraries, Grafana, Jupyter…; Tensorflow, Keras, FI-WARE Orion Context Broker, Apache Flink, Storm, Spark…) identified per analysed technology (data analytics/visualisation; big data, among others) (more information in Section 3) have been classified. Based on this classification framework, the matching between these tools and the i4Q solutions (more information in Section 4) has been performed collaboratively to provide a first overview of the i4Q solutions technological needs.

Similarly, task T1.1 has defined a classification framework that contains the following aspects: language/libraries, containers, orchestration; database; operating system (OP); user interface; synchronous interface; asynchronous interface; network; and other external solutions. The objective of this framework is to collect relevant information about the matching between the previous aspects and the i4Q solutions.

In order to be able to compare the different technologies identified, an evaluation benchmarking framework was defined. Section 2 describes this framework and explains thoroughly the dimensions and criteria that will be used for the evaluation of technologies and tools. Furthermore, the evaluation benchmarking framework facilitates the comparison among the different technologies based on the proposed set of criteria, which are also classified into several different dimensions. The comparison has been defined based on a '+' or '-' evaluation. Once that the different technologies and solutions were identified and the structure and contents of the framework defined, each technology was described. Technologies' description is developed in Section 3, including information regarding the state of the specific technology, its maturity, tools, EU project solutions, benchmarking and assessment, and application of i4Q solutions. A brief summary of the technologies has been provided in order to avoid making the deliverable too lengthy. However, each technology presents a set of references that describe in detail each of the identified technologies for further information.

Technologies and tools were characterized and evaluated based on the six above-mentioned dimensions and several evaluation criteria in order to investigate which technologies meet most requirements and seem to be more suitable, efficient and easier to be used during the development of i4Q solutions. Moreover, the 22 proposed i4Q solutions were also evaluated and characterized according to their matching and suitability with the specific technologies and tools. Results of these two analyses are presented in Section 4. Next, Section 5 addresses the application of the analyzed digital technologies and tools in the 22 proposed i4Q solutions. Finally, the deliverable ends with the conclusions in Section 6.

2. Benchmarking Evaluation Framework

The benchmarking evaluation framework is composed by different dimensions that, in turn are divided into different criteria to assess the different technologies that will be analysed and selected as relevant to fulfil the development objectives of the i4Q solutions.

Six dimensions have been defined to classify and aggregate the different criteria. The dimensions are the following ones:

- **General**. This dimension involves the generic criteria related to the general characteristics of each technology. This dimension includes aspects such as flexibility, specificity, performance, among others. Moreover, this dimension includes the criteria "relevance" that will be the one used as a summary of all the criteria of all the dimensions as a single quantitative measure.
- **Technological**. This dimension involves all the technical aspects for the evaluation of the technologies.
- **Business Model**. This dimension is one of the broadest as it involves aspects such as financials (costs of implementing and/or deploying the technology); regulatory (legislative factors that have to be taken into account as they hinder the use of technology); risks (factors that endanger the use or deployment of the technology, among others.
- **Informational**. This dimension is related to the information/data characteristics that the use of a specific technology could provide.
- **Social**. This dimension is related to aspects associated to the people who have to implement, deploy and / or use the technology/tools.

The following criteria have been defined to be part of the evaluation criteria (alphabetically ordered):

- **Capability/Features**. This criterion is related to the fitness and alignment of the technology with the i4Q solutions' needs and requirements.
- **Cost**. This criterion takes into account the cost of this technology and also the cost of using and integrating this technology into the development of the i4Q solutions. For this reason, open-source technologies will be encouraged. However, it is worth mentioning that although the technology is freely available, some technologies are more costly to implement than others due to a greater number of modifications and adjustments.
- **Coverage**. This criterion is related to the extent to which the technology covers the IT requirements.
- **Development-friendliness**. This aspect is related to how intuitive and easy-to-develop is the technology to be used for the i4Q solutions.
- **Genericity**. This criterion is related to the universality of the technology, that is the more general the technology is, the more concrete aspects will need to be designed but more freedom to personalize different functionalities needed in the i4Q solutions.
- **Integration**. This criterion is related to the ease of integrating the technology into the existing and/or legacy systems.
- **Interoperability**. This criterion is related to the level of information/data sharing/exchange that the selected technologies could provide.

- **Learning curve**. This criterion is related to how long it takes by developers to acquire new skills and/or knowledge to use the technology.
- **Legal Compliance**. This criterion is related to the regulatory aspects associated to the selected technology. The more complex and restrictive these regulatory aspects are, the more complex the adoption and use of such technology for the development of i4Q solutions.
- **Maturity**. This criterion is related to the level of the technology development/advance when the technology is released.
- **Need for data traceability**. This criterion is a specific one as it is related to the data management of the technology. i4Q solutions need to ensure the necessary data identification and recognition and the technologies selected need to guarantee this aspect.
- **Need for quality data**. Quality of data is a key factor for different technologies such as machine learning applications. It is crucial to have quality information of products and/or processes and this criterion is related to the capability of technologies to ensure the quality of data.
- **Performance**. This criterion is related to the fulfilment of the solutions development goals by the technology, i.e. the level with which the technology achieves the development requirements.
- **Relevance**. This criterion will assess the overall significance on the applicability of such technology in the development of the i4Q solutions. This criterion will take into account in a quantitative way, the assessment performed in the other factors.
- **Risk**. This criterion is related to the maturity of the selected technology and also to the security aspects. The less mature the technology, the riskier its selection will imply.
- **Scalability**. This criterion is related to the level with which the technology could increase its capacity and functionalities based on users' demand.
- **Security**. This criterion is related to the safety and protection that the selected technology could provide in the development, implementation and usage of the i4Q solutions.
- **Social preferences**. This criterion is related to the level of acceptance of developers/end users in regard to this technology use and adoption. It is not related to how much the technology is easy-to-understand and use but it is related to the prejudice against a specific technology and/or tool as the opinion or feeling held before careful analysing the technology.
- **Support**. This criterion is related to existence of a community-based support for such a technology.
- **Traceability**. This criterion is related to the ability of the technology to instantaneously track and trace the required aspects.
- **Training and documentation**. This criterion is related the learning curve, as the longer the learning curve is, the more training and documentation will be required.

Dimensions	Criteria	Technologies			
		Tech 1	Tech 2	Tech 3	Tech 4
General	Relevance				
	Capability/Features				
	Genericity				

Dimensions	Criteria	Technologies			
		Tech 1	Tech 2	Tech 3	Tech 4
Technological	Performance				
	Coverage				
	Interoperability				
	Integration				
	Scalability				
	Security				
	Traceability				
Business model	Cost				
	Maturity				
	Legal Compliance				
	Risk				
	Support				
Informational	Need for data traceability				
	Need for data quality				
Social	Development-friendliness				
	Learning curve				
	Social preferences				
	Training and documentation				

Table 1. Structure of the Benchmarking Evaluation Framework.

Once that the different set of criteria has been defined and classified into different dimensions (***Table 1***), each technology and associated tools will be analysed based on this evaluation framework. Due to the heterogeneousness of the different technologies analysed in Section 3, a rough evaluation including various criteria will be performed following the below mentioned principles:

- '+' symbol means a positive evaluation of a specific criteria.
- '-' symbol means negative evaluation of such criteria.
- 'NA' means 'not applicable' as such criteria does not have direct relationship with the technology or tool benchmarked.

The positive or negative evaluation is related to the fitness, appropriateness, and usefulness of the application of a technology in the i4Q project.

Moreover, in order to provide a more quantitative benchmarking, the analysis will use 1, 2 or 3 '+' or '-' symbols to quantify the degree of positiveness or negativity of such criteria in the assessed technology as it is shown in ***Table 2***.

+	++	+++
Slightly	Moderately	Very
-	--	---
Slightly little	Moderately little	Very little

Table 2. Meaning of the evaluation symbols.

For example, with the criteria of interoperability, '+' means that the technology is slightly interoperable, '++' means that the technology is moderately interoperable while '+++' means that the technology is very interoperable. Following the same example, '-' means that the technology is slightly little interoperable, '--' means that the technology is moderately little interoperable, '---' means that the technology is very little interoperable.

Moreover, it is important to clarify that some criteria have favourable connotation (the objective is to maximize such a criteria), while other criteria have an unfavourable connotation (minimize). ***Table 3*** defines, for each of the criteria, the connotation to avoid misunderstanding. It is worth mentioning that the '+' symbol will always mean a positive evaluation, whether it is intended to maximize or minimize, while the symbol '-' will always mean a negative evaluation.

Dimensions	Criteria	Connotation	Description
General	Relevance	Maximize	+++ It is very relevant. --- It is very little relevant.
	Capability/Features	Maximize	+++ It presents very relevant. --- It is very little relevant.
	Genericity	Maximize	(Freedom to customize functionalities) +++ It is very general. --- It is very little general.
	Performance	Maximize	+++ It has very high performance. --- It has very little performance.
	Coverage	Maximize	+++ It has very high coverage. --- It has very little coverage.
Technological	Interoperability	Maximize	+++ It is very interoperable. --- It is very little interoperable.
	Integration	Maximize	+++ It has very high integration. --- It has very little integration.
	Scalability	Maximize	+++ It is very scalable. --- It is very little scalable.
	Security	Maximize	+++ It is very secure. --- It is very little secure.
	Traceability	Maximize	+++ It is very traceable. --- It is very little traceable.
Business model	Cost	Minimize	+++ It is very costly. --- It is very little costly.
	Maturity	Maximize	+++ It is very mature. --- It is very little mature.
	Legal Compliance	Minimize	+++ It has very little legal regulations. --- It has very high legal regulations (increased complexity).
	Risk	Minimize	+++ It is very little risky. --- It is very risky.
	Support	Maximize	+++ It has very high support. --- It has very little support.
Informational	Need for data traceability	Maximize	+++ It is very data traceable. --- It is very little data traceable.
	Need for data quality	Maximize	+++ It manages data with very high quality. --- It manages data with very little quality.
Social	Development-friendliness	Maximize	+++ It is very development-friendly. --- It is very little development-friendly.
	Learning curve	Minimize	+++ It has a very little learning curve. --- It has a very high learning curve.
	Social preferences	Maximize	+++ It is very socially preferable. --- It is very little socially preferable.
	Training and documentation	Minimize	+++ It needs very little training and documentation.

Dimensions	Criteria	Connotation	Description
			--- It needs very high training and documentation.

Table 3. Description of the '+' and '-' symbols.

3. Digital Technologies for Manufacturing Quality and Data Reliability

3.1 Blockchain/ Hyperledger

3.1.1 Description

Blockchains are digital ledgers implemented without a central repository (distributed fashion) (Yaga et al., 2019). Blockchains give users the ability to record transactions in a shared ledger within a community, such that the transactions will not be changed after being published (Yaga et al., 2019). A blockchain consists of data sets which are composed of a chain of data packages, that are called blocks. Each block comprises multiple transactions. By adding more blocks, the blockchain is extended and thus represents a complete ledger of a transaction history (Nofer et al., 2017). The first block is called the "genesis block". Every block contains a timestamp, a unique hash value of the "parent", which is actually the previous block and a nonce, a random number for verifying that hash (Nofer et al., 2017). If a change has been detected in a hash value, that means that blocks have been changed, which would be a result of fraud. This way the frauds can be prevented immediately (Nofer et al., 2017). In order for a block to be added in the chain, it requires that the majority of nodes in the network agree by a consensus mechanism on the validity of transactions in a block and on the validity of the block itself.

In Blockchain applications, a shared ledger among users (organizations, businesses, individuals, software (SW) agents, etc.) enables users to agree on its content and secure all the transactions that cannot be modified after their addition, as mentioned earlier. In addition, the following actions are permitted: detailed tracking, measurement and tracing of transactions. The agreed activity can be a payment transaction between two members or a medical lab test entry for a patient (Al-Jaroodi and Mohamed, 2019).

The fundamental features of blockchain, as mentioned in (Al-Jaroodi and Mohamed, 2019), are the following:

- Digital identities: Blockchain can be considered as the digital equivalent of an identity card and thus can be used to authenticate people, organizations or any kind of entities. It can also be extended to include some more features, like property and possessions.
- Distributed security: blockchain can protect data and transactions recorded in the shared ledger by supporting different levels of encryption and hash features that users can employ. When a new transaction is added, it is firstly validated by the participating entities and then linked with the chain of previous transactions, so it is impossible for any records to be altered in any way.
- Smart contracts: that permit the conducting of a trackable, secure and unalterable contract over a public network without a third party. Blockchain-based smart contracts have the ability to advance many industrial sectors in different ways. like automating agreement processes between companies, partners and customers.

Micro-controls: securely recording events and activities without the need for third party confirmations and assurances. Blockchain technology is widely used in the financial sector in the form of crypto currency, like the popular Bitcoin and in other domains, especially for applications relevant to decentralized procedures and security (Nofer et al., 2017). In the industrial sector, blockchain technologies are used in applications and offer many advantages, like increased efficiency and security; enhanced traceability and transparency; and reduced costs (Al-Jaroodi and Mohamed, 2019).

Hyperledger is a joint project[1] of open source blockchains and other related tools, started in December 2015 by the Linux Foundation and has received contributions of well-known industries, like IBM, Intel and Systems, Applications, Products in Data Processing SAP[2]. The project was created with the aim to build an open-source distributed ledger framework and code base (Cachin, 2016).

In terms of architecture there is a consensus protocol that executes a replicated state machine that accepts three types of transactions as operations: deploy, invoke and query transaction. Validation of these transactions occurs through the replicated execution of the chaincode and given the fault assumption underlying the (BFT) consensus (Lachin, 2016).

Although blockchain applications in manufacturing are increasing, due to digital transformation, the successful industrial applications are still rare (Lohmer and Lasch, 2020). The same stands for relevant scientific publications. We list below some applications of blockchain in the manufacturing sector, found in the 2020's review (Lohmer and Lasch, 2020):

- developing a secure manufacturing platform to license products on a blockchain automatically with ensured data protection and traceability" (Bartsch et al., 2018)
- "blockchain testbed platform for sharing logs of machine events through smart contracts in a permissioned network." (Angrish et al., 2018)
- "use blockchain technologies for sharing and scheduling resources in distributed manufacturing networks." (Lohmer, 2019)

Characteristics	Description
Background	Distributed Digital ledgers.
Concept	A chain of blocks with multiple transactions that can be extended and provides high levels of safety and tracking of procedures.
Functions	Crypt wallets, Security.
Applications	• Financial applications, e.g., bitcoin. • Medical applications, e.g., health platforms, secure patient data entry and storage. • IoT applications: security of devices, marketplace, management of 'things'.
Effects	Security, safety, crypto wallets.
Key Technologies	Blockchain/ hyperledger.
Result verification	*Yes*

Table 4. Blockchain/Hyperledger at a glance

[1] https://www.hyperledger.org/

[2] source: wikipedia

3.1.2 State and Maturity

Blockchain technologies were introduced around 2015 and they are popular for creating crypto wallets. Blockchain provides a high level of security and thus it keeps becoming very popular in security related applications. The first know Blockchain application was the bitcoin. Hyperledger, that is a blockchain project, started off in December of 2015 by Linux foundation. In early 2016, it already had 60 members (Cachin, 2016). Both are quite widespread and enjoy increasing popularity.

Regarding the manufacturing sector, an evaluation of the application of blockchain can be found in (Sandner et al, 2017). Manufacturing companies have started to use blockchain solutions with 3D printing to enable new and safe manufacturing processes. Other applications include the supply chain for tracking items when shipped and a wide variety of applications in the IoT, like device identification over blockchain, timestamping of sensor data and marketplace, just to name a few.

3.1.3 EU Project Solutions

Table 5 shows the acronym of the project in which diverse blockchain/hyperledger solutions were developed (the acronym is a link to the project web page), the grant agreement reference (GA), the solution and its description.

Acronym	GA	Solution	Description
Blockpool	828888	Pooling SME (Small Medium Enterprise) adoption and deployment of Blockchain and other DLTs	The project employs Blockchain and Distributed Ledger Technologies (BDLTs) to enhance innovation capacity in enterprises. .
SOFIE	779984	Secure Open Federation for Internet Everywhere	The project used Distributed Ledger Technology (DLT) , including blockchains and inter-ledger technologies, in order to enable actuation, auditability, smart contracts and management of identities and encryption keys, as well as enable decentralised solutions with virtually unlimited scalability.
Bros	751615	Blockchain: a new framework for swarm RObotic Systems	A framework for robotic swarm systems with novel combination of blockchain technology , in order to generate new models that provide data confidentiality and entity validation to robot swarms.
BLOOMEN	762091	Blockchains in the new era of participatory media experience	A blockchain-enabled framework for creative content management.

Table 5. Blockchain/Hyperledger EU related-results

3.1.4 Tools

Table 6 presents a list of blockchain/hyperledger related solutions.

Solutions*	Description
Hyperledger Fabric	An implementation of a distributed ledger platform for running smart contracts, with a modular architecture that allows pluggable implementations of various functions.[3]
Solidity	A very popular language used by Blockchain Developers for writing smart contracts, that is influenced by C++, Python, and JavaScript. it was designed to target the Ethereum Virtual Machine (EVM). It is useful for writing applications that can execute self-enforcing business logic embodied in smart contracts Solidity has the following characteristics: • statically typed, • supports inheritance, libraries, and complex user-defined types. • supports the OOP paradigm and cis Available on: ethereum, ethereum classic, tendermint, counterparty, Tron, Hedera Hashgraph.[4]
Geth	Geth is an Ethereum node implementation that automatically connects to the Ethereum main net. It is available in three interfaces and can operate on all three major operating systems – Windows, Mac, and Linux. It is used as a host of different tasks on the Ethereum Blockchain, like transferring tokens, mining ether tokens, creating smart contracts, and exploring block history. Once Geth is installed, the user can either connect to an existing Blockchain or create their own.
Mist	Mist is the official Ethereum wallet. Before start using the Ethereum platform, a user must have a designated place to store Ether tokens and execute smart contracts. Mist is available for Windows, Mac and Linux.
Solc	Solc (Solidity Compiler) is a Solidity command-line compiler written in C++ that converts Solidity scripts into a more readable format for the Ethereum Virtual Machine. The smart contracts written in Solidity need to be converted to a format that can be easily read and decoded by the EVM and Solc is used for this reason.
Remix	Remix IDE is a Blockchain tool, written in Javascript and it is used for writing, testing, debugging and deploying of smart contracts, written in Splidity. It can be used locally or in the browser.
Metamask	Metamask is a wallet that acts as acts as a browser extension and a bridge between Ethereum Blockchain and a Chrome or Firefox browser. Once the app is installed in the browser, a built-in Ethereum wallet is ready to be used. It can interact with different Ethereum test networks.
Truffle	Truffle is an Ethereum Blockchain framework that is used to create a development environment for Ethereum-based apps. It has a large library of custom deployments for writing new smart contracts and develop complex Ethereum dApps. Truffle can perform automated contract testing using Chai and Mocha.
Ganache	Ganache is a Blockchain tool from the Truffle Suite that allows the user to create their own private Ethereum blockchain for testing applications (dApps) and executing commands among others. Its most important feature is the ability to perform all the actions that would otherwise be performed on the main chain. It is a useful tool for Blockchain Developers for testing their smart contracts during development.

[3] https://github.com/hyperledger/fabric

[4] Source: wikipedia

Solutions*	Description
[Blockchain Testnet]	A Blockchain Testnet is a very useful tool for testing dApps for bugs and errors, before making them live. Each blockchain solution has its unique Testnet, and there are three types of Blockchain Testnets – Public Test, Private Test, and GanacheCLI.

**The solutions' name is a link to the web page of the companies' solution developer.*

Table 6. Blockchain/Hyperledger Tools[5].

3.1.5 Benchmarking and Assessment

This section performs a global assessment of the blockchain/hyperledger technology in general (***Table 7***).

Dimensions	Criteria	Connotation	Description
General	Relevance	+++	
	Capability/Features	+++	
	Genericity	++	
	Performance	+++	
	Coverage		
Technological	Interoperability	++	
	Integration	+++	
	Scalability	++	
	Security	+++	
	Traceability		
Business model	Cost	++	Variety of open tools and scripts
	Maturity	+++	Not old technology, but has been already implemented a lot
	Legal Compliance	+	
	Risk	+	
	Support	++	
Informational	Need for data traceability	+++	
	Need for data quality	+++	
Social	Development-friendliness	++	
	Learning curve	++	
	Social preferences	++	
	Training and documentation	++	

Table 7. Blockchain/Hyperledger Benchmarking

3.1.6 Application for i4Q Solutions

Table 8 lists the i4Q project's solutions that could potentially use blockchain technologies.

Solution	Description
i4Q^BC	i4Q Blockchain Traceability of Data
i4Q^PQ	i4Q Data-driven Continuous Process Qualification

[5] https://www.upgrad.com/blog/top-blockchain-tools/

Table 8. Blockchain/Hyperledger Application for i4Q Solutions.

3.2 Fog/Edge Computing

3.2.1 Description

The development of the IoT has led to the need for processing data where it is produced due to the response delay that can be generated by the network. The data generated by users' devices are all transferred to distant clouds to be stored or processed. This computing model is not practical because it is likely to increase communication latencies when millions of devices are connected to Internet in the future (Hong, C. H., 2019). In many cases the systems require a low-latency interaction with its environment or users and this latency degrading the overall Quality-of-Service and Quality-of-experience. This is where the concept of the alternative Fog/Edge Computing model was born.

Edge computing is defined as computational processes that are performed inside edge devices (IoT devices) with analysis and processing capabilities such as compacts Pc, Raspberrys, routers or network gateways. By processing the obtained information close to where it was created, latencies are reduced and less bandwidth is consumed. However, the computing of this devices is limited in comparison with the Cloud.

Edge computing is characterized by low latency, dense geographical distribution, network context information, location awareness and proximity (Hassan N., 2018). The different objectives for edge computing in the context of IoT are as follows (Khan, W. Z., 2018):

Objectives	Description
Latency Minimization	Latency has become an increasing problem for IoT applications. Edge computing looks for minimizing it.
Cost Optimization	The use of an adequate platform for enabling edge computing requires an extensive infrastructure deployment. By setting up a number of nodes in a suitable position it can be reduced the cost from the network node placement.
Network Management	IoT needs an efficient usage from network resources in edge computing.
Energy Management	One of the principal objectives from edge system is to have a strict control over power management.
Resource Management	Optimal management of computational resources is crucial in obtaining service-level objectives.
Data Management	Efficient and effective data management mechanisms are desirable in edge computing.

Table 9. Edge Computing Objectives.

The concept of Fog Computing refers to a decentralized network structure in which resources (data, applications, etc.) are set in a logical place between the cloud and the generated data source (Gedeon J., 2019). It is to say, it brings the services that analyse such data closer to the source. Fog Computing emerges as an extension of Cloud computing (which consists of a distant server where data, applications, etc., are stored) to complement each other satisfying those needs between users' devices and Cloud Computing Centres.

Fog computing is characterized by low latency, interoperability, real-time interaction, location awareness and support for online interplay with the cloud. The different objectives for Fog computing in the context of IoT are as follows (Tadapaneni,2019):

Objectives	Description
Latency Minimization	The Fog server includes the localization of IoT devices to analyse and manage data. Therefore, having low latency it will work more efficiently and faster with real-time data.
Network Management	Fog computing devices are connectors between Cloud and IoT devices. The applications can process data when needed. This reduces data traffic sent to the cloud, eventually saving network bandwidth.
Resource-constrained devices	Fog computing helps the limited processing capabilities from IoT devices; therefore, it reduces power consumption and cost associated with the devices.
Uninterrupted computing services	Fog computing provides the user with continuous and uninterrupted services. This means that it can operate without being connected to the cloud, since it has its own servers that are located near to where the data is being collected.
Securing IoT	Many devices have limited security functions. Thanks to Fog computing, it can provide security updates to these devices. Furthermore, it can be used to monitor the security status. Privacy, Access Control, authentication, Trust or Data Protection these are just some of the security features added by fog computing (Alrawais A., 2017).

Table 10. Fog computing objectives.

Figure 1 illustrates its description. The cloud is at the top of the pyramid, where many servers are hosted. It offers computing capabilities or storage. Between Edge and Cloud computing, there is a Fog platform, which is composed by nodes in different areas. The end-user gadgets (laptops, smartwatch, etc.) can be found in the base, along with different IoT devices that can reach a Fog node with low latency (Confais, 2016).

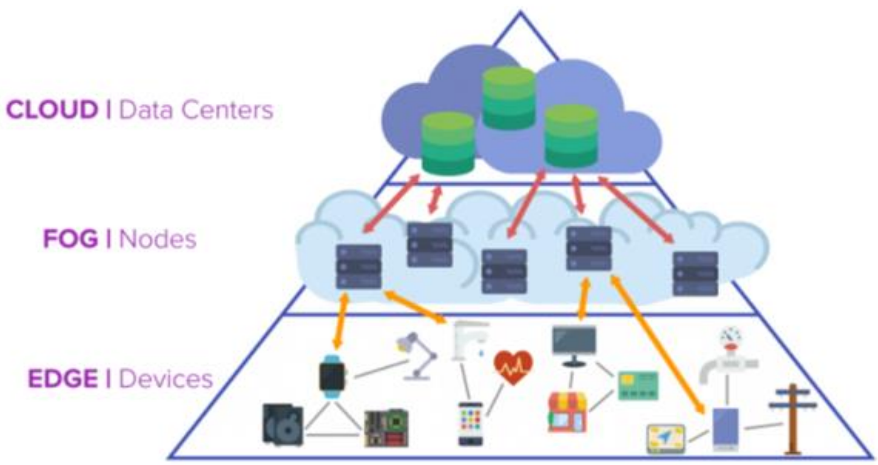

Figure 1. Fog/Edge computing (Erpinnews, 2018).

Characteristics	Description
Background	IoT and Cloud Computing.
Concept	Moving computational resources to or closer to data-generating devices reducing the impact on cloud.
Functions	Network architecture.
Applications	Wide variety of applications where use sensors and analyzing the data generated by IoT devices on the local network.
Effects	Reliability, low-latency, Geo-distribution, Mobily-support, scalability, interoperability, energy-efficiency and Bandwidth savings (Heck M., 2018).

Characteristics	Description
Key Technologies	There are several technologies that use this network architecture, such as OpenFog, IBM Edge Application Manager or ZDMP (Zero Defects Manufacturing Platform) Distributed Computing.

Table 11. Fog/Edge computing at a glance

3.2.2 State and Maturity

The IoT is accepted for any organization who wants to make their systems intelligent in industry 4.0 through any intelligent device, process control, supervisory control, and data acquisition (SCADA). This network structure has proven to be very beneficial for enterprises. Therefore, companies migrating to the use of IoT devices apply this structure. Also, large companies such as Amazon (AWS), IBM or Microsoft, offer services to integrate this model. Thanks to this, many companies have developed a platform adapting Edge/Fog computing concepts to different tools, systems, or technologies.

3.2.3 EU Project Solutions

Table 12 shows the acronym of the project in which this fog/edge computing solution was developed (the acronym is a link to the project web page), the grant agreement reference (GA), the solution and its description.

Acronym	GA	Solution	Description
5G-CORAL	761586	A 5G Convergent Virtualised Radio Access Network Living at the Edge	Deliverable that reports edge and fog computing in Tasks 2.1 and 2.2. Edge computing with 5G creates opportunities in industry. This project leverages the pervasiveness of edge and fog computing in the Radio Access Network (RAN) to create a unique opportunity for access.
FORA	764785	Fog Computing for Robotics and Industrial Automation	Deliverable that reports edge and fog computing in Tasks 1.1 D4. FORA focus on a reference system architecture for Fog Computing; resource management mechanisms and middleware for deploying mixed-criticality applications in Fog; safety and security assurance; service-oriented application modelling and real-time machine learning.
RECAP	732667	Reliable Capacity Provisioning and Enhanced Remediation for Distributed Cloud Applications	RECAP reports fog/edge computing in a case of use to demonstrate the capabilities of RECAP for automating the reallocation of resources closer to the edge network.
Fed4IoT	814918	Federating IoT and cloud infrastructures to provide scalable and interoperable Smart Cities applications, by introducing novel IoT virtualization technologies	This project faces the interoperability issue, focusing on scaling environments and addressing the problem at different levels.

Table 12. Fog/Edge computing EU related-results.

3.2.4 Tools

Table 13 presents a list of fog/edge computing related solutions.

Solutions	Description
AWS IoT Greengrass	Amazon Web Services (AWS) IoT Greengrass provides pre-built components so a user can easily extend edge device functionality without writing code. AWS IoT Greengrass components enable the user to add features, and quickly connect to AWS services or third-party applications at the edge.
AWS IoT SiteWise	AWS IoT SiteWise (Wireless Sensor Networks) is a managed service that makes it easy to collect, store, organize, and monitor data from industrial equipment at scale to help the user make better data-driven decisions.
IBM Edge Application Manager	International Business Machine (IBM) supply a solution for edge environments that helps the user to safely create, deploy, run, monitor, maintain and scale business logic and analytic applications at the edge. It can be run anywhere and it manages workloads on virtually any edge endpoint, including servers, gateways and devices.
ThingWorx Kepware Edge	ThingWorx Kepware Edge allows Kepware features to be deployed directly at the site of machines, devices, or sensors. This product can scale to create a distributed connectivity architecture.
Bosch IoT Edge	Bosch IoT Edge is an integrated set of tools and services that work together to connect different IoT devices locally and to the cloud, set communication between them, and develop scalable IoT applications that bring together IoT device data processing and services where they can best optimize outcomes.
Microsoft Azure IoT Edge	This is a fully managed service built on Azure IoT Hub. It provides a cloud workload to run on IoT edge devices via standards containers.
Oracle Roving Edge Infrastructure	Oracle roving edge infrastructure provides a deployment of cloud outside the datacentre. Thanks to this, cloud computing and storage services at the edge of networks will allow faster insights and processing closer to the data source.
FogAtlas	It is a framework that aims to manage geographically distributed and decentralised infrastructure that provides computing, storage and network services close to data sources and users, adopting the cloud computing paradigm. With this tool, the so-called Cloud-to-Thing Continuum can be managed and it also facilitates the operations of a Fog Computing infrastructure.
OpenVolcano	OpenVolcano aims to control and support 5G infrastructure to support mobile edge and fog computing services. It provides us with a modular architecture and a set of Application Programming Interfaces (APIS) that allows us to create Infrastructure as a Service (IaaS) and Platform as a Service (Paas) services.
KubeEdge	kubeEdge is based on Kubernetes and provides infrastructure support between cloud and edge. It is a native container orchestrator that manages devices to hosts on the edge. It also synchronises metadata between the cloud and the edge.
ZDMP Distributed Computing	Facilitates creation of logical structure for location of computational devices which can be used in efficient distribution of computational tasks at either Edge, Fog or Cloud level.
EdgeXFounDry	A Linux foundation common open platform for IoT edge computing. EdgeX Foundry focus is to exploit the benefits of Edge Compute by leveraging cloud-native principles and by enabling an architecture that meets specific needs of the IoT edge
AKRAINO	It is an open-source software stack to improve the state of edge cloud infrastructure. Edge computing applications running in virtual machines and containers.
StarlingX	Starlingx is an open stack project. Run the minimal services required at the edge, yet provide robust support for bare metal, container technologies and virtual machines.

Solutions	Description
OPNFV	Open Platform for Network Functions Virtualization (NFV) - (OPNFV) is a collaborative project under the Linux Foundation that is transforming global networks through open-source Network Functions Virtualization (NFV).
Eclipse Foundation	There are a several projects under Eclipse Foundation supports IoT and Edge platforms (IoT Cloud platform Stack, Eclipse Hono, Eclipse Kura).
ETSI	Multi-Access Edge Computing (MEC) offers application developers and content providers cloud-computing capabilities and an IT service environment at the edge of the network. This environment is characterized by ultra-low latency and high bandwidth as well as real-time access to radio network information that can be leveraged by applications.
Open Connectivity Foundation	The Open Connectivity Foundation is dedicated to ensuring secure interoperability for consumers, business and industries by delivering a standard communications platform, a bridging specification, an open-source implementation and a certification program allowing devices to communicate regardless of form factor, operating system, service provider transport technology or ecosystem.
ITU	Specifies signalling requirements and architecture of intelligent edge computing to provide intelligence to the edge network for efficient data processing within the network.
OpenFog	To drive industry and academic leadership in fog computing & networking architecture, testbed development, and interoperability and composability deliverables that seamlessly bridge the cloud-to-things continuum.
Open Horizon	Open Horizon is a platform for managing the service software lifecycle of containerized workloads and related machine learning assets. It enables autonomous management of applications deployed to distributed webscale fleets of edge computing nodes and devices without requiring on-premise administrators.
Nerve Blue	Nerve Blue is a radically open edge computing platform that promotes vendor independence and flexibility. Nerve Blue runs on off-the-shelf hardware, scaling from gateways to Inter-Process Communications (IPCs). Thanks to its open architecture users can deploy their own legacy software, or applications developed and provided by third parties. An intuitive management system that is located at the edge or in the cloud allows user to manage all available nodes that are registered to the system.

Table 13. Fog/Edge Computing Tools.

3.2.5 Benchmarking and Assessment

This section performs a global assessment of the fog/edge computing technology in general (***Table 14***).

Dimensions	Criteria	Connotation	Description
General	Relevance	+++	It is a community-accepted network architecture to reduce latencies and improve data processing with IoT devices.
	Capability/Features	+++	
	Genericity	+++	
	Performance	+++	
	Coverage	+++	Covers network architectures
Technological	Interoperability	+++	
	Integration	++	
	Scalability	+++	

Dimensions	Criteria	Connotation	Description
	Security	+	Administrators must implement security measures at the network level and always be up to date.
	Traceability	+	
Business model	Cost	++	
	Maturity	+++	
	Legal Compliance	+	
	Risk	++	Network Security risk
	Support	+	
Informational	Need for data traceability	+	
	Need for data quality	+	
Social	Development-friendliness	+++	
	Learning curve	--	
	Social preferences	+++	
	Training and documentation	--	

Table 14. Fog/Edge Computing Benchmarking.

3.2.6 Application for i4Q Solutions

Table 15 lists the i4Q project's solutions that could potentially use blockchain technologies.

i4Q Solution	Description
i4Q EW	i4Q Edge Workloads Placement and Deployment
i4Q AI	i4Q Artificial Intelligence (AI) Models Distribution to the Edge

Table 15. Fog/Edge Computing Application for i4Q Solutions.

3.3 Data Analytics

3.3.1 Description

Data analytics is a general term that encompasses multiple tasks relevant to data preparation and information extraction. It is usually distinguished from the term data science; however, this distinction is not always clear. Data analysis is usually the first step of extraction information from data, while data science is often used to define the more advanced analysis of predictive algorithms and exploration of more complex relations among data. A common data analysis pipeline is the following:

- **Data collection or extraction**: during the first stage the data that need to be processed are collected. Depending on the nature of the work, the data can be collected from databases (e.g., MongoDB), extracted from APIs, or they can be provided to the analyst in an editable format, like an excel or text file. Usually, the data are collected according to some requirements. These requirements could be which variables need to be stored, the type of variables (numeric, string, categorical), probably an indicative variable, like an id number etc.
- **Data processing/ cleaning**: after collecting or extracting the data, they need to be processed so that mistakes or redundant information will not be passed on the main analysis step. Data cleaning involves removal of duplicate registrations, removal of errors and removal or replacement of missing values.
- **Data filtering**: in certain types of data, like time series, it is useful to apply filtering techniques in order to reduce the noise. Some filtering techniques are moving averages, Kalman filtering and normalization.
- **Information extraction**: in this step, the data is ready to be further analysed in order to derive some basic useful information. In the data analytics level, the information that is usually needed includes: visualizations of certain variables or subsets of data, some basic relations among data, like correlations or distribution of numeric variables across different categories of other variables. Some of these techniques are often referred as exploratory data analysis, because they explore the relationships among data (Schutt and O'Neil, 2013).

Data analytics are applied in multiple scientific or business applications that involve collection and exploitation of data.

Characteristics	Description
Background	Statistical and probability theory.
Concept	To form data into an exploitable format and extract information from data.
Functions	• Data cleaning. • Data pre-processing. • Relations among data (e.g., correlations, chi-square tests). • Visualizations.
Applications	Wide variety of applications, including all fields and tasks that exploit data: • Social sciences (e.g., demographics, polls). • Medical sciences (e.g., clinical data analysis). • Financial applications. • IT. • Construction and manufacturing.
Key Technologies	Statistical software (e.g., SPSS (Statistical Package for the Social Sciences), STATA, Minitab) and programming languages (e.g., R, Python).
Data Sources	This is not a limited field. Data can be derived from many sources. Some of the most common are:

Characteristics	Description
	• Databases (e.g., Mongo D.B). • Readable data files [e.g., Excel, Text (txt)]. • Cloud. • Social media (e.g., text data from Twitter). • API's.
Data Volume	Data analytics can be applied on large datasets, however when the size and complexity increase, big data techniques are used.
Data acquisition	The data can be either provided as a separate file or extracted by a database or by calling an API.
Data processing	• Filtering: removes noise from data. • Missing values imputation: replace missing values of a variable with the variable's average or using a predicted value. • Removal of duplicates and errors.
Visualization	• Visualization of frequency distributions: histograms, line charts, pies. • Visualization of relations among data: scatterplots, heatmaps, boxplots.

Table 16. Data Analytics at a glance

3.3.2 State and Maturity

Data analytics is a wide term that encompasses many methodologies and procedures. It is hard to define a specific time of appearance, since there are different processes included in the term, however data analytics is a field quite old, e.g., Pearson's correlation coefficient has been introduced since 1920s, and widely exploited. Data analytics have already reached a high level of maturity and the constantly increasing need for data driven solutions, may lead to further advancements in tools and algorithms.

Regarding the maturity of data analytics in the manufacturing sector, an empirical study of 2017, that researched 100 manufacturing companies, located mainly in Germany and Switzerland, found that in most companies the amount of data that was exploited, was quite smaller that the available amount of data. The problem for this focuses mainly on the processing step of the data and on the information extraction (Groggert et al., 2017).

3.3.3 EU Project Solutions

Table 17 shows the acronym of the project in which diverse data analytics solutions were developed (the acronym is a link to the project web page), the grant agreement reference (GA), the solution and its description.

Acronym	GA	Solution	Description
INNOVATE	745829	INtelligeNt ApplicatiOns oVer Large ScAle DaTa StrEams	A set of solutions based on large scale data analytics for the management of continuous data streams

ELASTIC	825473	A Software Architecture for Extreme-ScaLe Big-Data AnalyticS in Fog CompuTIng ECosystems	Two deliverables that explain the data analytics used in the project's solution: "D7.2 Data Management Plan" and "D2.5 Distributed Data Analytic Platform".
ZDMP	825631	Zero Defect Manufacturing Platform	A platform for eliminating defects in manufacturing, that employs data analytics techniques. A deliverable about artificial intelligence and analytics techniques "D074 - AI and Analytics".
BOOST 4.0	780732	Big Data Value Spaces for COmpetitiveness of European COnnected Smart FacTories 4.0	Data analytics solutions for manufacturing data. A deliverable "D3.3 Big Data Models and Analytics Platform", where besides the state of the Art, the analytic services of the platform are described.

Table 17. Data Analytics EU related-results

3.3.4 Tools

There are numerous software solutions for data analytics. Excel is probably the oldest one and traditional statistical software like SPSS, Stata, Weka, as well as programming languages like R, Python and SAS. There are also solutions for data visualization like Tableau. In the following table though, only the open access tools are mentioned, since those will be used during the project.

Solutions	Description
Microsoft azure	Azure is a platform with many products relevant to data analytics and artificial intelligence. Some of these programs are provided for free, while others require a fee. The platform offers a complete set of solutions and capabilities regarding data analytics.
Python libraries	Python is a programming language. The most well-known libraries for data analytics are: NumPy, SciPy, Pandas, Seaborn
R libraries	R is a statistical programming language. Most of the basic data analytics tasks can be conducted with the base package. Other packages relevant to data analytics are: Dplyr, Ggplot2, Shiny,
Grafana	Grafana is an analytics platform that offers multiple visualization and alerting tools.
Apache superset	Apache superset is a lightweight platform for data exploration and visualization

Table 18. Data Analytics tools.

3.3.5 Benchmarking and Assessment

This section performs a global assessment of the data analytics technology in general (***Table 19***).

Dimensions	Criteria	Connotation	Description
General	Relevance	+++	
	Capability/Features	+++	
	Genericity	+++	
	Performance	+++	
	Coverage	+++	
Technological	Interoperability	+++	

Dimensions	Criteria	Connotation	Description
	Integration	+++	
	Scalability	++	
	Security	++	It depends on other security protocols
	Traceability	+++	
Business model	Cost	++	
	Maturity	+++	
	Legal Compliance	+++	
	Risk	++	Network Security risk
	Support	+++	
Informational	Need for data traceability	+++	
	Need for data quality	+++	
Social	Development-friendliness	+++	
	Learning curve	+	
	Social preferences	+++	
	Training and documentation	+++	

Table 19. Data Analytics Benchmarking.

3.3.6 Application for i4Q Solutions

Table 20 lists the i4Q project solutions that could potentially use data analytics technologies.

i4Q Solution	Description
$i4Q^{IM}$	i4Q Infrastructure monitoring
$i4Q^{DIT}$	i4Q Data Integration and Transformation Services
$i4Q^{DA}$	i4Q Services for Data Analytics
$i4Q^{AD}$	i4Q Analytics Dashboard
$i4Q^{DT}$	i4Q Digital Twin simulation services
$i4Q^{PQ}$	i4Q Data-driven Continuous Process Qualification
$i4Q^{PA}$	i4Q Prescriptive Analysis Tools
$i4Q^{LRT}$	i4Q Manufacturing Line Reconfiguration Toolkit
$i4Q^{LRG}$	i4Q Manufacturing Line Reconfiguration Guidelines

Table 20. Data Analytics Application for i4Q Solutions.

3.4 Big Data

3.4.1 Description

The emergence of technologies and the realization of the value of data has led to the big data era, where huge amounts of data are stored in order to be used for the extraction of useful information. With the term “big data” we refer to datasets with great volume and velocity that are also usually quite unstructured[6]. Another characteristic is that these amounts of data need to be processed quite fast, e.g., to produce results real time. These characteristics cause issues in the storing, handling and visualization of data. The HACE theorem describes the characteristics of big data (Wu et al., 2013):

- **H**uge data with Heterogeneous and Diverse Dimensionality: The heterogeneity in big data comes for the fact that there are multiple data sources with different protocols and structures.
- **A**utonomous Sources with Distributed and Decentralized Control: The autonomous data sources generate information without centralized control, which adds to the heterogeneity issue of Big Data.
- **C**omplex and **E**volving Relationships: Data relations become more complex as the size of the dataset increases. Thus, it is required that there are certain variables that characterize the dataset, e.g., demographic characteristics of individuals. Later on, the complex relations among data will be filtered by the application of dimensionality reduction techniques and feature selection methods that aim to remove redundant information from data.

Characteristics	Description
Background	Data analytics methodologies, database handling, multisource data integration.
Concept	To be able to handle big volumes of complex data in terms of memory and performance.
Functions	• Handling of large volumes of data. • Functions used in data analytics and machine learning, adjusted for the amount, volume and complexity of data.
Applications	• Computer vision: medical applications, security applications. • IoT systems. • Systems biology, genomics, metabolomics, etc.
Key Technologies	• Apache hadoop/ apache spark. • Keras. • Tensorflow.
Data Sources	This is not a limited field. Data can be derived from many sources. Some of the most common are: • Databases (e.g., Mongo D.B). • Readable data files (e.g., Excel, txt). • Cloud. • Social media (e.g., text data from Twitter). • API's.
Data Volume	The size of a dataset in order to be considered big data is not strictly defined. It could be a combination of number of rows and columns, e.g., a

[6] https://www.sas.com/el_gr/home.html?gclid=CjwKCAjwgZuDBhBTEiwAXNofRMbXWCiI6VACQ0-IB8QLkg-KEao1W28J5GFcWsvlVokuhHiGW25ewRoCv3UQAvD_BwE

	dataset of hundreds of thousands of cases and 200 variables, can be considered big data.
Data acquisition	The data can be either provided as a separate file or extracted by a database or by calling an API.
Data processing	Yes. All typical functions used in data analytics, with emphasis on dimension reduction techniques.
Visualization	Visualization options for exploring and observing the data
Result verification	The results can be verified with the use of various metrics, like accuracy, sensitivity, etc.

Table 21. Big Data at a glance.

3.4.2 State and Maturity

The term big data has been used since 1990s, but the exploitation of big data is continuously expanding and has been on the rise for the last 10 years. Currently there are many tools to handle and analyze big data and address the challenges that arise in storing, analysis and interpretation.

3.4.3 EU Project Solutions

Table 22 shows the acronym of the project in which diverse big data solutions were developed (the acronym is a link to the project web page), the grant agreement reference (GA), the solution and its description.

Acronym	GA	Solution	Description
LAMBDA	809965	Learning, applying, multiplying big data analytics	A deliverable 2.1 "Big Data Challenges and Analysis of Scientific and Technological Landscape" with the current state of the art tools relevant to Big Data applications.
BigDataStack	779747	High-performance data-centric stack for big data applications and operations	A complete solution for infrastructure management that utilizes big data analytics.
NOESIS	769980	Novel decision support tool	Deliverables that outline the uses of big data in transport (D2.1, D2.2, D3.3).
Monsoon	723650	MOdel based coNtrol framework for Site-wide OptmizatiON of data-intensive processes	A framework with big data analytics services. ff Deliverables focused on big data: D4.3 - Initial Big Data Storage and Analytics Platform; D4.4 - Updated Big Data Storage and Analytics Platform; D4.5 - Final Big Data Storage and Analytics Platform.

Table 22. Big Data EU related-results

3.4.4 Tools

Table 23 presents a list of machine learning related solutions.

Solutions	Description
Tensorflow	TensorFlow is an open-source platform for machine learning. It provides a flexible, easy to use ecosystem of tools, libraries and community resources that allows scientists to apply state-of-the-art ML methods and deploy their own applications.
Keras	Keras is a library of open-source software that is widely used for modelling neural networks
Apache Kafka	Kafka is an event streaming platform, for read/write streams of events, store and processing of streams.
Apache Storm	Apache Storm is an open source distributed real time computation system. that allows the processing of streaming data. It can be used with any language and is widely known for the analysis of sensor recordings.

Table 23. Big Data Tools.

3.4.5 Benchmarking and Assessment

This section performs a global assessment of the big data technology in general (***Table 24***).

Dimensions	Criteria	Connotation	Description
General	Relevance	+++	
	Capability/Features	+++	
	Genericity	++	
	Performance	+++	
	Coverage	+++	
Technological	Interoperability	+++	
	Integration	+++	
	Scalability	++	
	Security	++	
	Traceability		
Business model	Cost	++	
	Maturity	+++	
	Legal Compliance	+	
	Risk	++	Network Security risk
	Support	+++	
Informational	Need for data traceability	++	
	Need for data quality	+++	
Social	Development-friendliness	++	
	Learning curve	+++	
	Social preferences	+++	
	Training and documentation	+++	

Table 24. Big Data Benchmarking.

3.4.6 Application for i4Q Solutions

Table 25 lists the i4Q project solutions that could potentially use big data technologies.

i4Q Solution	Description
i4Q^{IM}	i4Q Infrastructure monitoring
i4Q^{DIT}	i4Q Data Integration and Transformation Services
i4Q^{DA}	i4Q Services for Data Analytics

i4Q^{BDA}	i4QBig Data Analytics Suite
i4Q^{PQ}	i4Q Data-driven Continuous Process Qualification
i4Q^{QD}	i4Q Rapid Quality Diagnosis

Table 25. Big Data Application for i4Q Solutions.

3.5 Machine Learning

3.5.1 Description

The term includes a number of statistical algorithms used for prediction and classification. Machine learning is considered a combination of statistics and informatics. There are two categories, the supervised learning and unsupervised learning. Supervised learning includes methodologies that train algorithms on labelled data. Unsupervised learning refers to clustering, where the groups in the data are not known beforehand. The following terms are important to understand machine learning applications:

- **Independent variables**: the variables that their values do not depend on other variables. They will be used as an input in a predictive algorithm.
- **Dependent or target variables**: the variable that needs to be predicted. Upon the type of the target variable, i.e., whether it is numeric or categorical, depends on the choice of the prediction algorithm.
- **(True) labels**: the levels of a categorical target variable are called labels. The true labels are fundamental for supervised learning, since there is the need to know the categories of the data in order to train a model.

A pipeline for machine learning includes the following steps:

- **Data cleaning/ preparation**: the first step before conducting any analysis. It is described in Data analytics section.
- **Feature extraction**: this is optional. It is more often met in applications of time series data, where features need to be computed from the original values in order to eliminate noise.
- **Feature selection**: Very important step that removes redundant features and improves the results of the prediction or classification algorithm.
- **Sub-setting of data**: the data are split into two or three groups: training and test set or training, validation and test set respectively. The training set is used to train the algorithm, so that it can 'learn' from the data which values correspond to which level of the target variable. The validation set can be utilized to fine tune the parameters of an algorithm, which translates to running an algorithm multiple times, with different values of the algorithm's parameters and decide on which combination produces the optimal results. The validation set is not found in all applications. The test set is used for prediction and evaluation of results.
- **Application of algorithm**: if the target variable is categorical, a classification algorithm is applied, while for continuous target variables, regression algorithms are used. Examples of classification algorithms are: Support vector classification, random forests, logistic regression. Regression algorithms are linear regression, support vector regression.
- **Evaluation of results**: if there are available labels in the test set, the predicted results can be cross-checked against the true labels and thus evaluate the performance of the

prediction. For the evaluation the following metrics are used: accuracy, f-score, sensitivity/specificity, roc curves.

Deep learning is a “branch” of machine learning that is used in computer vision problems, natural language processing and audio recognition among others. The term "deep" refers to the use of multiple layers in the network. The term includes the following architectures: deep neural networks, deep belief networks, artificial neural networks and other modifications[7]. Deep learning models represent multiple processing layers and more complex relations between the data. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep neural networks have achieved remarkable results in problems where conventional machine-learning techniques were not able to handle. Neural networks are fed with raw data and are able to automatically discover the relations needed for detection or classification (LeCun, Bengio and Hinton, 2015).

Characteristics	Description
Background	Machine learning is considered a field in between statistical science and informatics.
Functions	Prediction and classification are the main functions and final targets of supervised machine learning algorithms. These algorithms can be applied in both real-time and non real-time applications. For the unsupervised type of machine learning, the main function is to discover homogeneous groups of data.
Applications	Wide area of applications, in all fields that data can be collected: • Social sciences, demographics. • Medical, biological science. • Business processes, finance. • Industrial applications.
Key Technologies	• Supervised learning: Support Vector Machines, Random Forests, Decision trees, Regression etc • Unsupervised learning: k-means clustering, hierarchical clustering etc
Data Sources	This is not a limited field. Data can be derived from many sources. Some of the most common are: • Databases (e.g., Mongo D.B). • Readable data files (e.g., Excel, txt). • Cloud. • Social media (e.g., text data from Twitter). • API’s.
Data Volume	Depending on the algorithm and the number of variables, a minimum number of cases is usually requested. Machine learning algorithms can handle big amounts of data, however for large datasets big data methodologies and deep learning are preferred.

[7] Source: wikipedia

Characteristics	Description
Data acquisition	The data can be either provided as a separate file or extracted by a database or by calling an API.
Data processing	The data processing before entering an algorithm can include the following, besides cleaning and preparation: • Feature extraction. • Feature selection. • Transformation (e.g., From a categorical to numeric dummies).
Visualization	Various visualization tools available, like: • Scatterplots. • Boxplot. • Regression errors' plots.
Result verification	For the evaluation of results there are certain metrics like: • Accuracy. • Precision. • F-score, etc.

Table 26. Machine Learning at a glance.

3.5.2 State and Maturity

Machine learning has been quite popular and continuously expanding in the last 20 years. One can generally argue that it is a quite mature field, that allows for experimentation and expansion of the algorithms.

3.5.3 EU Project Solutions

Table 27 shows the acronym of the project in which diverse machine learning solutions were developed (the acronym is a link to the project web page), the grant agreement reference (GA), the solution and its description.

Acronym	GA	Solution	Description
PROTEUS	687691	Scalable online machine learning for predictive analytics and real-time interactive visualization	The project developed a library of Scalable Online Machine Learning and Data Mining Algorithms (SOLMA) that includes distributed online algorithms for clustering, classification, regression etc. There is available code and deliverable that describes the solution: "D4.3 Novel Scalable Online Machine learning algorithms for data streams".
MUSKETEER	824988	Machine learning to augment shared knowledge in federated privacy-preserving scenarios	The project provided secure and scalable machine learning solutions over decentralized datasets. Deliverables describing machine learning algorithms for federated operation modes (D4.4) and for semi-honest operation modes (D4.6).
AutoML	899880	Unlocking the potential of machine learning for SMEs by automated machine learning	The project will develop an affordable, automated machine learning (AutoML) method to enable efficient implementing of most advanced ML applications on industrial settings.

SLING	81978	Efficient algorithms for sustainable machine learning	A solution of resource-efficient algorithms for large-scale machine learning.

Table 27. Machine Learning EU related-results.

3.5.4 Tools

The traditional statistical software can be used for machine learning applications, like SPSS, R, PYTHON and WEKA. However, programming languages allow more freedom when building and assessing the models.

Solutions	Description
R	R is a language and environment for statistical programming and visualizations. R provides a wide variety of libraries for statistical and graphical applications, and is highly extensible.
PYTHON	Python is a programming language that also offers multiple libraries for statistical analysis and also allows for integration of systems more effectively. .
Apache Spark MLib	MLib id apache's libraby for machine learning algorithms
Mahout	Mahout is an Apache framework suitable for the application of custom machine learning algorithms.

Table 28. Machine Learning Tools.

3.5.5 Benchmarking and Assessment

This section performs a global assessment of the machine learning technology in general (***Table 29***).

Dimensions	Criteria	Connotation	Description
General	Relevance	+++	
	Capability/Features	+++	
	Genericity	+++	
	Performance	+++	
	Coverage	+++	
Technological	Interoperability	+++	
	Integration	+++	
	Scalability	++	
	Security	+	
	Traceability	+++	
Business model	Cost	++	
	Maturity	+++	
	Legal Compliance	+	
	Risk	++	Network Security risk
	Support	+++	
Informational	Need for data traceability	+++	
	Need for data quality	+++	
Social	Development-friendliness	+++	
	Learning curve	++	

	Social preferences	+++	
	Training and documentation	++	

Table 29. Machine Learning Benchmarking.

3.5.6 Application for i4Q Solutions

Table 30 lists the i4Q project solutions that could potentially use machine learning technologies.

i4Q Solution	Description
$i4Q^{IM}$	i4Q Infrastructure monitoring
$i4Q^{DIT}$	i4Q Data Integration and Transformation Services
$i4Q^{DA}$	i4Q Services for Data Analytics
$i4Q^{BDA}$	i4QBig Data Analytics Suite
$i4Q^{DT}$	i4Q Digital Twin simulation services
$i4Q^{PQ}$	i4Q Data-driven Continuous Process Qualification
$i4Q^{QD}$	i4Q Rapid Quality Diagnosis
$i4Q^{PA}$	i4Q Prescriptive Analysis Tools

Table 30. Machine Learning Application for i4Q Solutions.

3.6 IIoT

3.6.1 Description

IIoT aims at connecting industrial assets, like engines, power grids and sensor to cloud over a network (Helmiö, 2018).

IIoT involves several devices connected by communications software. The IIoT systems, considering the different devices that are part of such systems are in charge of monitoring, collecting, exchanging, analysing, and instantly acting on information to smartly adapt their behaviour or their environment without human interaction (RTI, 2015).

Another perspective (Emmrich et al., 2015) involves that IIoT is not just based on Cyber-Physical System (CPS), but also on embedded systems, cloud computing, edge computing, the generic technologies associated with the smart factory, and associated software.

IIoT is a system that comprises networked smart objects, CPS assets, connected generic information technologies and optional cloud or edge computing platforms, which allow real-time, smart, and autonomous access, collection, analysis, communications, and exchange of process, product and/or service information, within the industrial environment, with the main objective to optimise the whole production value. This can be achieved by improving product or service distribution, enhancing productivity, decreasing labour costs, energy consumption, and reducing the products' manufacturing cycle (Boyes et al., 2018).

Characteristics	Description
Background	An evolution of the programmable logic controller (PLC).
Concept	A system involving networked smart objects, which allow real-time, smart, and autonomous access, collection, analysis, communications, and

Characteristics	Description
	exchange of process, product and/or service information, with optimisation purposes.
Functions	Real-time, smart, and autonomous access, collection, analysis, communications, and exchange of process, product and/or service information.
Applications	• Autonomous vehicles. • Machine utilisation. • Operator productivity. • Power management. • Quality control. • Smart logistics. • Wearables.
Effects	• Enable flexible and agile production. • Shorten lead times with reliable wireless devices. • Improve operational performance. • Guarantee safety and compliance.
Key Technologies	• Hardware (Arduino, raspberry...). • Software, operating systems (Contiki, OpenWrt, ...). • Frameworks (AllJoyn, Countly...). • Middleware (Kaa, Losant...).
Data Sources	Data from physical devices (sensors...).
Data Volume	No specific quantity, but high quantity.
Data features	Often unstructured data.
Multisource correlation	Focus on the consistency of multi-source data and their evolution and integration.
Data acquisition	Sensors.
Data processing	No specific methods.
Visualization	Data interfaces.

Table 31. IIoT at a glance.

3.6.2 State and Maturity

IIoT deployments are evolving quickly with regards to the maturity chain, progressing from pilot projects to large and real scale implementations providing great benefits (Kennan, 2020).

The Bsquare 2017 Annual IIoT Maturity Survey results show that 86% of the surveyed participants have adopted IIoT solutions and 84% believe their solutions are very or extremely effective (Bsquare, 2017).

Surveyed participants from manufacturing industries revealed their use of the IIoT adoption stages:

- Device's connectivity and simple data forwarding: 67%
- Real-time dashboards and monitored activity through constant data stream: 62%
- Advanced analytics, such as machine learning, cluster analysis, and artificial intelligence, from data scientists: 47%
- Automated simple, single-step actions such as service-ticket requests: 32%

IIoT return on investment raises when companies use advanced analytics (e.g., machine learning), for identifying data correlations in real-time, and for the processes' automation (e.g. repair tickets). Because the use of these two stages lags behind device connectivity and real-time monitoring, manufacturing companies have significant opportunity to advance their IIoT capabilities for greater future impact. 90% of the surveyed respondents affirm that IIoT will have either a significant or tremendous global impact on their industry.

3.6.3 EU Project Solutions

Table 32 shows the acronym of the project in which IIoT solution was developed (the acronym is a link to the project web page), the grant agreement reference (GA), the solution and its description. As designing and implementing a IIoT system requires the simultaneous adoption of several technologies and tools, when the search for IIoT, different technologies are found.

Acronym	GA	Solution	Description
Andromeda	783664	System uses AI and Industrial IoT to usher in new era of predictive maintenance in rail	System that integrates smart IIoT devices and Artificial Intelligence (AI) into the very first end-to-end predictive maintenance solution for railway infrastructure.
SERENA	767561	Predictive AI condition-based maintenance solutions for advanced manufacturing	A scalable and resilient IIoT platform enabling Artificial Intelligence services that aim to reduce costs and improve productivity.
ZDMP	825631	D059 - Data Acquisition and IIoT	Deliverable that describes the Data Acquisition as a prerequisite for data access, exchange, and analytics within ZDMP component. It implements a framework for the handling of data from IoT sensors and other sources. Provides the second version of the applications' AI methods and solutions.
SEMIoTICS	780315	Embedded Intelligence and local analytics	This deliverable focuses on the building blocks required for introducing embedded intelligence to IIoT / IoT things and especially to local analytics mechanisms, so as to enable semi-autonomic local reaction / adaptation within IIoT / IoT devices.

		Bootstrapping and interfacing SEMIoTICS field level devices	This deliverable provides the design and first draft of semantic specification of SEMIoTICS field level devices, as well as their use for bootstrapping and easier integration of the devices in IIoT systems.
AirWatt	732493	Coin-sized microturbine powers smart gas grid solutions	AirWatt offers reliable power for off-grid remote solutions.

Table 32. IIoT EU related-results.

3.6.4 Tools

In the Gartner's June 2019 Magic Quadrant for IIoT Platforms, there are no leaders or challengers. Moreover, its 2018 report, showed that there was no "dominant provider" for IIoT platforms. Kosan (2019) defines, according to industry analysts, an IIoT platform as a platform that should provide:

- Device management software that connects physical and digital systems.
- Integration through software development kits, development tools and APIs to support business processes and enterprise systems across the business.
- Data management to control and monitor ingestion, storage, accessibility, and flow.
- Analytics of data from connected devices, the enterprise and third parties to reveal patterns and optimization of assets.
- Allow to configure and operate connected assets that enable digital twins.
- Software to enable security audits and guarantee compliance, such as mechanisms to avoid data loss and to identify and take action on breaches.
- Support for protocols used in the industrial domain.
- Engineering-level robustness to prevent downtown.
- Flexibility with no-code interfaces.
- A combination of cloud computing, on-premises deployment and edge computing.

Based on the analysis performed by Kosan, (2019) which is in turn based on Forrester (2019) and Gartner (2019), ***Table 33*** shows 10 IIoT platforms that fulfil the previous characteristics.

Solutions	Description
Atos: Atos Codex IoT	Atos is an established IIoT partner for utility companies seeking new opportunities (such as analytics in IoT), it plans to leverage its partnership with Siemens (see below) to address manufacturing and transportation. Atos focuses mainly on Siemens solutions, specifically integration with the MindSphere IoT platform.
Hitachi: Lumada	The Lumada platform launched in May 2016 and is available on premises or in the cloud. It offers different types of IoT-enabled cloud managed services, particularly to heavy industrials, according to the company. Lumada provides the full range of IIoT platform functionality, enhanced by Hitachi technologies and analytics and data management capabilities. The majority of deployments focus on improving operations and predictive maintenance. However, according to Gartner, Hitachi lacks a "broad and robust" IoT edge device management capability and is still making improvements in that area.
IBM: Watson IoT Platform	The Watson IoT platform, generally available since 2014, is a set of managed cloud services with the IBM Cloud. It can also be deployed on premises via private cloud.

Solutions	Description
	Watson IoT provides a "packaged functionality" from which customers can compose an IIoT solution, according to the company. IBM highlights its analytics and AI capabilities as key differentiators for Watson IoT.
Microsoft: Azure IoT	Azure IoT provides a set of development and advanced analytics tools. Microsoft emphasizes its end-to-end security capabilities open-source tools. It tends to support its own portfolio of products and should offer a more comprehensive set of pre-built applications, according to Forrester's assessment.
Oracle: Oracle IoT Cloud, Oracle IoT Applications	Oracle IoT Cloud Service envisions IIoT as an "IoT-enriched SaaS business application sale" rather than a PaaS technology sale, according to the company. The business applications focus is on monitoring IoT assets, production, fleets and services, and the connected worker. IIoT solutions need Oracle Cloud to be run however an on-premises "rendition" on the Oracle private cloud is available.
Schneider Electric: EcoStruxure	Schneider Electric's EcoStruxure IoT offering, available in early 2017 and built on top of Microsoft Azure IoT Suite, contains a set of scalable services and IIoT platform modules that are essential for industrial enterprises, according to the company. On-premises, edge and cloud solutions based on the EcoStruxure IoT platform are deployed across different industries.
Siemens: MindSphere	Siemens offers a set of devices for deployment at the edge, including the MindConnect Nano and Simatic IoT family. These are capable devices, but Siemens' edge connectivity strategy can look more dependent on the vendor's own hardware.
Software AG: Cumulocity IoT	Cumulocity, Software AG's foundational IIoT application suite platform, comprises Cumulocity IoT Core and Cumulocity IoT Edge and provides device administration (including device health and connectivity management) and preconfigured IIoT applications. Optional components can handle enterprise and cloud integration, real-time analytics, and data and API management. The platform can be deployed in the cloud, on premises and at the edge.

Table 33. IIoT Platforms.

Open-source tools

The open-source tools related to IIoT are classified into:

Hardware

- Arduino – It consists of an open-source electronics platform that provides user-friendly hardware and software.
- BeagleBoard – It is a low-power open-source hardware single-board computer developed by Texas Instruments in association with Digi-Key and Newark element14.
- Microduino – Microduino and mCookie offer potent, small, stackable electronic hardware.
- Node MCU (ESP 8266) – Node Microcontroller Unit (NodeMCU) is an open-source IoT platform. It uses the Lua scripting language. It is based on the eLua project, and built on the ESP8266 Software Development Kit, SDK 0.9.5.
- OLinuXino – OLinuXino is an open-source software and hardware low cost (approx. 30€) linux Industrial grade single board computer with General Purpose Inputs/Outputs (GPIOs).
- Raspberry Pi – It is a linux-based development board.
- Tessel – It is an open-source and community-driven IoT and robotics development platform. It involves development boards, hardware module add-ons, and the associated.

Software, operating systems

- Contiki – It is an open-source operating system for the IoT. It connects tiny low-cost, low-power microcontrollers to the Internet.
- OpenWrt – It is an embedded operating system based on the Linux kernel, primarily used on embedded devices to route network traffic. The main components are the Linux kernel, util-linux, uClibc or musl, and BusyBox.
- Tiny OS – It is an open-source, Berkeley Software Distribution (BSD)-licensed operating system for low-power wireless devices, e.g. those used in sensor networks, ubiquitous computing, personal area networks, smart buildings, and smart meters.
- Linux - It is an open-source operating system compatible with virtualisation technologies such as VirtualBox, Openstack or VMWare.

Frameworks

- AllJoyn – It is an open-source software framework that allows devices and apps to discover and communicate with each other.
- Countly IoT Analytics – It is a general-purpose and open-source analytics platform for mobile and IoT devices.
- Iotivity – It is an open-source software framework that allows the connectivity of device-to-device.
- Kura – It offers a Java/ Open Services Gateway initiative (OSGi) based container for Machine-to-Machine (M2M) applications running in service gateways. Kura provides or, when available, aggregates open-source implementations for the most common services needed by M2M applications.
- OpenHAB – The openHAB runtime is a set of OSGi bundles deployed on an OSGi framework (Equinox). It is a Java solution and needs a JVM to run. As it is based on OSGi, it offers a greatly modular architecture, which at the same time enables inserting and deleting functionality while the service is running.

Middleware

- Kaa –It is an open-source middleware platform for the creation of IoT solutions.
- Losant – It uses open communication standards like Representational state transfer (REST) and Message Queuing Telemetry Transport (MQTT) to provide connectivity among devices. It provides data collection, aggregation, and visualization features. Its drag-and-drop workflow editor allows to trigger actions, notifications, and M2M communication without programming.
- DreamFactory – It is a free open-source REST API platform for mobile, web and IoT applications.

3.6.5 Benchmarking and Assessment

This section performs a global assessment of IIoT technology in general (***Table 34***).

Dimensions	Criteria	IIoT	Description
General	Relevance	+++	
	Capability/Features	+++	
	Genericity	++	
	Performance	++	

Dimensions	Criteria	IIoT	Description
	Coverage	+++	
Technological	Interoperability	++	Proper connectivity is critical for interoperability: MQTT 5 protocol. Open-source IIoT solutions tend to provide enhanced interoperability and robust APIs and SDKs to be integrated with other systems and products.
	Integration	++	Having success in the integration of IoT into legacy industrial systems is challenging.
	Scalability	-	Although most of the industrial enterprises have implemented IIot solutions, only 5% can be able to scale their applications (BusinessWire, 2020).
	Security	-	Interconnected devices usually use ad-hoc protocols or gateways to get to universal protocols such as OLE for Process Control (OPC) Unified Architecture. One of the main concerns of IIoT are these custom protocols or gateways that often do not have enough security aspects (IIoT-World, 2021).
	Traceability	++	
Business model	Cost	++	
	Maturity	+++	
	Legal Compliance	++	
	Risk	-	Cybersecurity compromise in industrial IoT (device hijacking, data siphoning, denial of service attacks, data breaches, device theft, man-in-the-middle or device “spoofing”) (Gatto, 2021).
	Support	++	
Informational	Need for data traceability	+	
	Need for data quality	+	It is needed a stable and strong data network.
Social	Development-friendliness	++	
	Learning curve	++	
	Social preferences	++	
	Training and documentation	++	

Table 34. IIoT Benchmarking.

3.6.6 Application for i4Q Solutions

Table 35 lists the i4Q project solutions that could potentially use IIoT technologies.

i4Q Solution	Description
i4Q^{BC}	i4Q Blockchain Traceability of Data
i4Q^{TN}	i4Q Trusted Networks with Wireless & Wired Industrial Interfaces

Table 35. IIoT Application for i4Q Solutions.

3.7 Digital Twin

3.7.1 Description

Based on Haag and Anderl (2018) research, there are many meanings and definitions of the concept of digital twin. It seems that this term was firstly coined in the aerospace industry (Negri et al., 2017) with a focus on structural mechanics, material science and the long-term performance prediction of air and space crafts (Shafto et al., 2012; Tuegel et al., 2011). Currently with the new paradigm of Industrie 4.0, digital twin is focused on the support of ensuring information continuity throughout the entire product lifecycle (Abramovici et al., 2016; Rosen et al., 2015), virtual commissioning of (manufacturing) systems (Schluse and Rossmann, 2016), and decision support and system behavior predictions in the product development process and the next lifecycle phases based on computer-aided simulations (Kraft, 2016).

Digital twin is a collection of all digital artifacts that accumulate during product development process linked with all data that is generated during product use (Haag and Anderl, 2018). In light of this, the digital twin is a complete digital representation of an individual product. It involves the properties, condition and behavior of the real-life object through models and data. The digital twin is developed alongside its physical twin and remains its virtual counterpart across the entire product lifecycle.

The digital twin is being used in the whole product lifecycle management (PLM), from design, to manufacturing to service and operations (Steer, 2018). Currently, PLM is very time consuming with regards to efficiency, manufacturing, intelligence, service phases and sustainability in product design. A digital twin can join the product physical and virtual space (Tao et al., 2017). The digital twin allows enterprises to get a digital footprint of all of their products in all their life cycle phases (Parrot and Warshaw, 2017; Porter and Heppelman, 2015). In the manufacturing process, the digital twin represents a virtual replica of the near-time occurrences in the factory. To do so, multiple devices such as sensors are placed in the shop-floor to collect data from different domains, such as environmental aspects, behavioural features of the machine and work that is being performed. All this data is continuously communicating and collected by the digital twin (Parrot and Warshaw, 2017).

As aforementioned, the digital twin is a very abstract concept, but it is the digital representation of all the elements of a factory to be able to anticipate and plan machine / equipment shutdowns, to perform simulations, to improve product design characteristics, to reduce products design time, among others. All this is much easier if objects (factory, products…) are virtualized through digital twins. This will allow to have a digital representation of the entire production, design and maintenance management environment, to mention a few. To do so, it is necessary to evolve from traditional sensors, which basically was devoted to sensorizing a variable, to smart devices that provide information not only about the variable to be measured but also about its environment.

Digital twins serve many purposes; there is no single digital twin. Different stakeholders interact with digital twins from different perspectives and derive different types of value (Robins, 2020). Following the fact that there is not a commonly accepted definition, we would also like to highlight the following ones (Scully, 2021).

- **Siemens**, for example, defines Digital Twins as "a virtual representation of a physical product or process, used to understand and predict the physical counterpart's performance characteristics".
- **Dassault Systems** takes a narrower view and defines Digital Twin technology as "a virtual representation of what has been produced."
- **IBM** limits the scope to real-time data and defines it as "a virtual representation of a physical object or system across its lifecycle, using real-time data to enable understanding, learning and reasoning".

The commonality in most definitions is that Digital Twins are a "digital abstraction or representation of a physical system's attributes and/or behavior".

Characteristics	Description
Background	The rapid development of new IT and the desire for cyber-physical integration.
Concept	Digital abstraction or representation of a physical system's attributes and/or behaviour.
Functions	• Virtual authentication. • Simulation analysis. • High-reliability real-time control and monitoring. • Forecasting and diagnosis the problems. • Optimizing and enhancing the process.
Applications	Representation of all the product lifecycle from design to maintenance, repair & overhaul and associated manufacturing processes.
Effects	• Enhancement of efficiency. • High precision in the monitoring and management of different aspects. • Reduce cost and development cycle. • Support smart manufacturing.
Key Technologies	• IoT. • Data analysis. • Virtual reality. • Augmented reality. • CPS. • Simulation.
Data Sources	The data from physical entities and virtual models.
Data Volume	No specific data volume.
Data features	Structured, semi-structured and unstructured data.
Multisource correlation	Focus on the consistency of multi-source data and their evolution and integration.
Data acquisition	Any type of sensing devices.
Data processing	No specific methods

Characteristics	Description
Visualization	Image, video, virtual and augmented reality, ...
Result verification	• Through its own virtual simulation and evolution functions to execute pre-verification. • Relatively advance.

Table 36. Digital Twin at a glance.

Finally, it is worth mentioning the evolution of digital twins, from concept to standalone operational technology (OT) application to integration with enterprise IT and eventually to a future point where the digital twin either ceases to be recognised as an entity in itself, or perhaps more radically, models the entire enterprise. This evolution is shown in **Figure 2**.

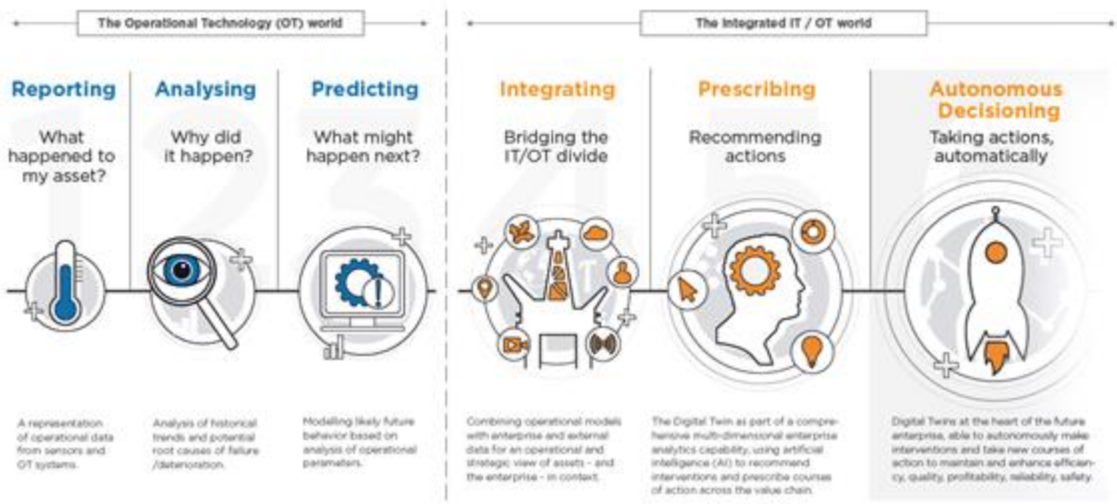

Figure 2. The Digital Twin Evolution (Source: Socha, 2018).

3.7.2 State and Maturity

The state and maturity of digitals twins have been analyzed from two perspectives: (i) from the technology viewpoint but also (ii) from the enterprises' adaptation capacity to this technology. This second perspective is very important to be taken into account as it is aligned with the work performed in T1.3 and T1.4.

With regard to the technology maturity level, no quantitative assessments have been found, only experts' judgements. The webcast, "Digital Twin: Unlocking IX in Process Plant and Industry hosted by LNS Research revealed the state of the technology and market and showed how companies can use it to drive operational excellence. The answer to the following question: "*Do you think that the technology (platforms) to implement digital twins is already mature enough (i.e. keeping pace with what can be done, and what customers would like to have, etc.)?*" points out that customers cannot yet know exactly what their needs are and what they would like to represent and for this reason, vendors are not sure about digital twin functionalities (Perino, 2019).

With regard to the enterprises' adaptation capacity to digital twins, it is worth mentioning that as digital twin is a replica of a physical thing - a 'twin', depending on maturity, this replica can range from a simple 2D or 3D model of a local component, all the way to a fully integrated and highly accurate model of an entire asset, facility or even a country, with each component dynamically linked to engineering, construction, and operational data (Tune, 2021).

Table 37 shows the digital twin maturity spectrum in which each of the elements further enables removing humans from hazardous processes or tasks, intrinsically improving safety.

Maturity Element (logarithmic scale)	Defining principle	Outline usage
0	Reality capture (e.g. point cloud, drones, photogrammetry)	• Brownfield (existing) as-built survey
1	2D map/system or 3D model [e.g. object-based, with no metadata or Building Information Modeling (BIM)]	• Design/asset optimisation and coordination
2	Connect model to persistent (static) data, metadata and BIM Stage 2 (e.g. documents, drawings, asset management systems)	• 4D / 5D simulation • Design / asset management • BIM Stage 2
3	Enrich with real-time data (e.g. from IoT sensors)	• Operational efficiency
4	Two-way data integration & interaction	• Remote & immersive operations • Control the physical from the digital
5	Autonomous operations & maintenance	• Complete autonomous operations & maintenance

Table 37. Digital twin maturity spectrum defining principles and outline usage (Tune, 2021)

Although organisations want to achieve the higher-order Elements 3 and 4, the reality is that most are only ready for Elements 0, 1 and 2. This should not be of concern, as each milestone provides incremental value (Tune, 2021).

Another maturity model that presents the different levels or stages of maturity of digital twins may be found in Robins (2020).

3.7.3 EU Project Solutions

Table 38 shows the acronym of the project in which this digital twin solution was developed (the acronym is a link to the project web page), the grant agreement reference (GA), the solution and its description. As designing and implementing a Digital Twin system requires the simultaneous adoption of several technologies and tools, when the search for digital twin, different technologies are found.

Acronym	GA	Solution	Description
SHERLOCK	820689	Workspace monitoring and digital twin – First prototype	Deliverable that reports the prototypes of the environment perception and digital twin models developed in Tasks 4.2 and 4.3

Acronym	GA	Solution	Description
PortForward	769267	Description of Port Virtual Twin Concept	Deliverable that establishes the digital twin approach for port environments based on the PortForward approach of using interoperable IoT applications and services
DECENTER	815141	First and Second release of applications' Artificial Intelligence methods and solutions	Deliverable that provides the second version of the applications' AI methods and solutions.
ScalABLE4.0	723658	A scalable solution for the factories of the future	A flexible, automated and scalable production system will allow manufacturers to quickly react to changing production demands.
SIMFAL	737881	Virtual and augmented reality tools for faster, cost-efficient aircraft assembly processes	Many tasks like aircraft cabin and cargo assembly are mainly performed manually under non-ergonomic conditions with complex process chains. To optimise all this, this solution has been developed.
MANUWORK	723711	Fostering human-machine collaboration in the factories of the future	This result is related to a computational method to optimise human-automation levels for workload balancing; a framework for the evaluation of worker satisfaction, safety and health; a solution for Industrial Social Networking in manufacturing; and another solution created to facilitate the use of Augmented Reality (AR) in the manufacturing workplace.
MooringSense	851703	Simulation Dataset	A first dataset generated through simulations with the available coupled numerical model will be provided to start the development of the Structural Health Monitoring (SHM) solution and the smart sensor.
DFD	889126	If a tree falls in the forest, these drones will spot it	A fleet of drones that can map every tree in a forest will make commercial forestry more powerful, more autonomous and more sustainable.

Table 38. Digital Twin EU related-results

3.7.4 Tools

Commercial Tools

Based on G2 (2021), digital twin software provides a virtual representation or simulation of a physical asset and its main objective is to examine the performance of the asset in real-time. These tools are used to simulate performance, predict potential maintenance needs, and ultimately optimize the asset. Businesses use sensors to produce the data necessary from their physical assets to provide the input data necessary for the digital twin. These IoT devices will allow enterprises to monitor and control the assets. These tools are often used in conjunction with IoT device management software or computer-aided engineering (CAE) software.

G2 (2021) classifies a product withing the digital twin category if the following conditions are fulfilled:

- Provide a digital representation of a physical asset.
- Track data produced by the physical asset using sensors.
- Allow enterprises to optimize performance of physical assets based on the data collected from the asset.

Based on the previous definition of digital twin software, **Table 39** presents a list of the solutions that fit with the two perspectives described in the previous definition. However, it is important to highlight that these tools are mainly focused on virtual representation or on simulation and only a few cover the broad scope of digital twin nature.

Solutions	Description
SAP Leonardo Internet of Things	This solution allows to: (i) unlock the smart enterprise by embedding IoT-enabled devices data into the core processes; (ii) optimize efficiency and productivity, redefine the customer experience, and reimagine business models to satisfy almost any customer's needs and fulfil all the business objectives.
Predix	Predix, the operating system for the Industrial Internet, is powering digital industrial businesses that drive the global economy.
aPriori Digital Manufacturing Simulation Software	aPriori is a provider of digital manufacturing simulation software. By leveraging the digital twin with the digital factories, aPriori automatically generates Design for Manufacturability (DFM) & Design to Cost (DTC) insights, helping manufacturers collaborate across the product development process to make better design, sourcing, and manufacturing decisions and create higher value products in less time.
IoTIFY	IOTIFY is a virtualization platform for the IoT.
Oracle IoT Production Monitoring Cloud	Oracle IoT Production Monitoring Cloud gathers data at the manufacturing machine level and based on production line and factory definitions, provides a holistic view of operations.
Akselos	Akselos is the creator of the advanced digital twin technology. Its products are developed particularly to support protect the world critical infrastructure with next-generation and advanced simulation technology.
Ansys Twin Builder	Ansys Twin Builder enhances predictive maintenance outcomes to save on warranty and insurance costs and optimize product's operations.
Autodesk Digital Twin	Autodesk helps companies to develop intelligent, data-rich digital prototypes of physical assets for both products and facilities. Therefore, it offers a platform for packaging digital twin BIM for handoff to owner/operators.
Bosch IoT Suite	Bosch IoT Suite provides the foundation for service enablement, both in terms of connecting things to the Internet reliably, securely, cost effectively and at scale and in terms of delivering the backing application logic for value-added services.
cohesion	Cohesion provides a building experience app that combines building systems, gives smart controls, and provides actionable insights addressed to commercial real estate. Although i4Q is not addressed to this sector, this technology has been included just for general interest.

Solutions	Description
CONTACT Elements for IoT	CONTACT Elements for IoT connect physical devices to the digital world. Its open architecture has been devised from the ground up to serve high-quality, critical assets in the most reliable way.
Flutura Decision Science	Flutura Decision Sciences and Analytics is an IoT intelligence company that is powering new monetizable business models using machine signals in the engineering and energy industry.
iLens	iLens from Knowledge Lens is an industrial IoT solution that provides, interface connectivity, edge computing, monitoring and control, and predictive analytics. iLens is currently be used by enterprises of different sectors such as automation, manufacturing, energy and utility.
Iotics	Iotics is a platform to surface meaningful, relevant and real-time insights from the entire data landscape — without duplicating, centralizing or re-architecting.
Machstatz	MachStatz. This solution uses artificial intelligence algorithms and machine learning diagnostics to provide advance analytical solutions. It uses smarter devices, sensors and gateway's which communicate with each other through LoRa, Narrow Band IoT (NB-IoT), etc. and other wireless technologies. The end-to-end solution provides the last mile connectivity for the customers where their need does not depend on other Original Equipment Manufacturers (OEM's) and vendors.
nD	nDimensional's product, nD, is a full stack application development Platform-as-a-Service (PaaS) that allows enterprise to quickly design, deploy and operate AI, big data and IoT applications. nD empowers data-rich industries to drive real-time actions that quantitatively enhance business outcomes, proven at enterprise scale, to bring the power of predictive insights and optimization to all vertical markets and value chains.

Table 39. Digital Twin Tools.

As it has been shown in **Table 39**, the commercial arena shows divergence. Based on Roest (2019) article; General Electric (GE Digital, 2021), Siemens (Siemens, 2021) and Dassault (Dassault Systèmes, 2018) each offer their own proprietary version of the common digital twin functionality. Microsoft (Mircrosoft, 2021) also has a digital twin but for now it seems to focus mostly on just collecting data for further analysis. The digital twins offered by Akselos (Akselos 2021)_perform structural simulations in near real time for the offshore industry but they seem less suited as a digital twin for production lines in factories. So, every player in the field puts forward its own interpretation of the concept.

Open-Source Tools

If the commercial arena presents many divergences related to the different scopes and tools that digital twins embrace, the open-source one also requires a common framework to work on so that researchers, software companies and end users can immediately make use of any progress that is made. This is essentially the call made by Cor van Kruijsdijk in his contribution to the ERCIM news issue titled (Roes, 2019). "Digital Twins as R&D Accelerators - The Case for an Open-Source Ecosystem" (Van Kruijsdijk, 2018).

From the open-source paradigm, the following initiatives have been identified:

Ditto project. (Eclipse, 2021)

This initiative has been developed by Bosch at creating an open-source foundational layer for digital twins. Eclipse Ditto is a technology in the IoT developing a software pattern called "digital twins". With Ditto a thing can just be used as any other web service via its digital twin.

Companies using Ditto will be able to get a fully-fledged, authorization aware API (HTTP, WebSocket and other messaging protocols) for interacting with digital twins and all aspects around them. For example, a company that is building an IoT solution using both hardware (e.g., sensors or actuators) and software (e.g., a mobile or web app) in order to solve a customer's problem. In such a scenario the company has several places where to implement software:

- on or near the hardware, e.g., on an Arduino using C/C++ or on a Raspberry PI using Python,
- optionally on a gateway establishing the Internet connectivity (e.g., based on Eclipse Kura),
- in the mobile or web app using Java, Javascript, Swift, etc.,
- in the "back end" fulfilling several responsibilities like
    - providing an API abstracting from the hardware,
    - routing requests between hardware and customer apps,
    - ensuring only authorized access,
    - persisting last reported state of hardware as cache and for providing the data when hardware is currently not connected,
    - notifying interested parties (e.g., other back-end services) about changes,

Ditto focuses on solving the responsibilities a typical "back end" has in such scenarios.

iModel.js. (iTwinsjs, 2021).

The other open-source digital twin project is iModel.js. It is a commercial initiative, this one connected to the US infrastructure company Bentley (Roes, 2019). It allows to create, visualize, and analyze infrastructure digital twins. It is open, flexible and built for the cloud, web and mobile context. It provides an open platform to build upon, to connect, and to be able to derive value from infrastructure digital twins across all enterprises' workflows. The system is written using modern cloud and web technologies and is available for building enterprises own solutions to visualize and connect with the rest of the digital world (CheManager, 2020).

The main objectives of iModel.js are described as follows (iTwinsjs, 2021):

- Be the open platform for Infrastructure Digital Twins for federating BIM Models, Geographic Information System (GIS) data, reality data, IT and OT into a single pane of glass for a seamless user experience.
- Be cloud provider, operating system, and browser neutral to the extent possible.
- Offer an improved performance, scalability, security, and stability
- To provide an easy, productive, predictable, and profitable platform.
- Follow open-source standards and norms wherever possible.
- Build an ecosystem of innovation.

From a more technical point of view, writing an iModel.js application requires the following software tools: (i) Node.js (12.x LTS version) - This provides the backend JavaScript runtime; (ii) Git - This is the source code control system for the iModel.js repositories; and (iii) Visual Studio Code - This is the recommended editor and debugger for iModel.js applications.

PyFMI / Functional Mockup Interface (FMI) compliant simulation Tool (i.e., OpenModelica)

PyFMI is a package that allows working with Functional Mock-up Units (FMUs) via an interface. It extends the FMI standard by providing a master code for orchestration of the co-simulation. Since the FMI standard does not bring a master code for orchestrating the co-simulation the PyFMI claims to deliver one at the latest state of art of co-simulation (Winther, 2020).

PyFMI is available as a stand-alone package or as part of the JModelica.org distribution. Using PyFMI together with the Python simulation package Assimulo adds industrial grade simulation capabilities of FMUs to Python. With Assimulo the solving of ordinary differential equations containing various solvers becomes possible (Andersson, 2016).

3.7.5 Benchmarking and Assessment

This section performs a global assessment of the digital twin technology in general (**Table 40**).

Dimensions	Criteria	Digital Twins	Comments
General	Relevance	++	
	Capability/Features	++	
	Genericity	-	
	Performance	+	
	Coverage	+++	
Technological	Interoperability	-	The lack of interoperability between the digital twins of different companies hinders use cases that require information exchange between different organization (Platenius-Mohr, 2019).
	Integration	-	
	Scalability	++	
	Security	-	The massive data utilized in the field of digital twin is prone to severe security breaches (Susila, 2020).
	Traceability		
Business model	Cost	--	A Digital Twin for a Grade a Commercial Office building of around 600,000 ft2 (60,000 m2) would cost $1.2m and $1.7m. Whereas, for a

Dimensions	Criteria	Digital Twins	Comments
			more complex and larger General Hospital building of around 2,100,000 ft2 (200,000m2) would cost $3m and $4.2m (Lengthorn, 2021).
	Maturity	-	
	Legal Compliance	++	
	Risk	+	
	Support	+	
Informational	Need for data traceability	+	Engineering data originates in diverse, multidiscipline applications and requires a great deal of aligning, validating and scrubbing which can turn out to be as laborious as it sounds. The golden rule is to ensure that data is accessible; in other words, do not get locked in by proprietary formats (Bentley, 2020).
	Need for data quality	-	
Social	Development-friendliness	+	
	Learning curve	+	
	Social preferences	-	
	Training and documentation	++	

Table 40. Digital Twin Benchmarking.

3.7.6 Application for i4Q Solutions

Table 41 lists the i4Q project solutions that could potentially use digital twin technologies.

i4Q Solution	Description
i4Q^{DT}	i4Q Digital Twin simulation services.
i4Q^{PQ}	i4Q Data-driven Continuous Process Qualification.

Table 41. Digital Twin Application for i4Q Solutions.

3.8 Security Technologies

3.8.1 Description

The security technologies have been classified into three main categories:

Standards, particularly 62443 standard (ISA, 2021)

The ISA/IEC 62443 series of standards, developed by the ISA99 committee and adopted by the International Electrotechnical Commission (IEC), is addressed to mitigate past, present and future security threats with regards to the industrial systems. These standards could be applied to all the industrial sectors and critical infrastructure.

ISA-62443-4-2, *Security for Industrial Automation and Control Systems: Technical Security Requirements for IACS Components*, provides the cybersecurity technical requirements for components that make up an industrial and automation control system (IACS), particularly the embedded devices, network and host components, and software applications. This standard is based on the ISA/IEC 62443-3-3, *System Security Requirements and Security Levels*. It specifies security capabilities that enable a component to mitigate vulnerabilities for a given security level without the assistance of compensating countermeasures.

Digital identity life cycle (asymmetric cryptography)

A Digital Identity management solution creates, manages and revokes digital identities. By having these identities, it is possible to enforce authentication, create access management policies and provide accountability. Identity management, therefore, involves a broader concept, which is a technology-neutral approach to integration and a flexible architecture that allows interoperability with several identity systems within and outside enterprises (AET, 2021).

Hardware root of trust (RAMBUS, 2021)

A hardware root of trust is the basis on which all secure operations of a computing system depend. It contains the keys used for cryptographic functions and enables a secure boot process. It is inherently trusted, and therefore must be secure by design. The most secure implementation of a root of trust is in hardware making it immune from malware attacks. As such, it can be a stand-alone security module or implemented as security module within a processor or system on chip (SoC).

A silicon-based hardware root of trust falls into two categories: fixed function and programmable. Essentially, a fixed-function root of trust is a state machine. These are typically quite simple, small and designed to perform a specific set of limited functions like data encryption, certificate validation and basic key management. However, these functions are static as a fixed function hardware root of trust can only do what it is specifically designed to do.

In contrast to its fixed-function counterpart, a hardware-based programmable root of trust is built around a Central Processing Unit (CPU). Performing all the functions of a state machine-based solution, a programmable root of trust can also execute a more complex set of security functions. A programmable root of trust is versatile and upgradable, enabling it to run entirely new cryptographic algorithms and secure applications to meet evolving attack vectors.

Characteristics	Description
Concept	Security technologies are addressed to mitigate past, present and future security threats from a hardware and software perspective.
Functions	• Data Loss Prevention. • Intrusion Detection/Prevention System. • Event Management. • Firewall protection. • Access control.
Applications	• Privacy Engineering. • Safety Communications. • Distributed Denial of Service (DDoS) protection. • Web application firewall. • Advanced Bot Protection.
Effects	• To protect critical data. • To ensure protecting all the physical assets. • To guarantee confidentiality, integrity, and availability.

Characteristics	Description
Key Technologies	• Standards. • Digital identity life cycle (asymmetric cryptography). • Hardware root of trust.
Data Sources	The data from physical entities and software applications.
Data Volume	No specific data volume, although it is expected that petabytes of data per second could be generated from normal activities.
Data features	Data diversity and disorganization.
Data acquisition	No specific method to acquire data from multiple sources.
Data processing	Data is converted into actionable information.
Visualization	Any type.
Result verification	Result verification allows enterprises to leverage threat intelligence to evaluate the effectiveness of controls against particular attacks.

Table 42. Security Technologies at a glance.

3.8.2 State and Maturity

62443 standard

In order to present the maturity of the standard 62443, it is important to know the evolution from ISA99 to International Electrotechnical Commission IEC 62443. **Figure 3** shows such an evolution.

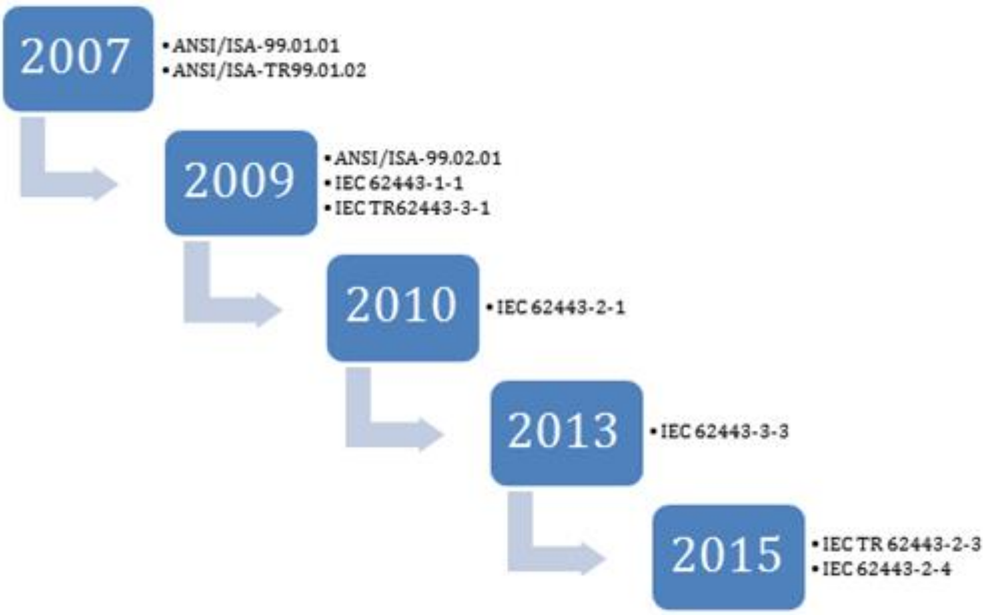

Figure 3. Evolution from ISA99 to IEC 62443 (source: INCIBE, 2015).

The IEC 62443 standard is composed by the following documents together with their state of maturity:

- IEC 62443-1-1 “Models and Concepts”: Published.
- IEC TR 62443-1-2 “Master Glossary of Terms and Abbreviations”: Published.
- IEC 62443-1-3 “System Security Compliance Metrics”: Draft version.
- IEC TR 62443-1-4 “Security Life Cycle and Use Cases”: Proposed for approval, but not approved yet.
- IEC 62443-2-1 “Requirements for an IACS Security Management System”: It is being reviewed to align its contents with International Organization for Standardization ISO 27000.
- IEC TR62443-2-2 “Operating a Control Systems Security Program”: Under development.
- IEC TR 62443-2-3 “Patch Management in the IACS Environment”: Under development.
- IEC 62443-2-4 “Certification of IACS supplier security policies and practices”: Under development.
- IEC TR62443-3-1 “Security Technologies for IACS”: In the review phase.
- IEC 62443-3-2 “Security Risk Assessment and System Design”: Published.
- IEC 62443-3-3 “System Security Requirements and Security Levels”: Published.
- IEC 62443-4-1 “Product Development Requirements”: Published.
- IEC 62443-4-2 “Technical Security Requirements for IACS Components”: Non-published.

As it can be seen in the different documents that make up the standard, most of them are still in the development, draft or/and under review phase. Therefore, the maturity state of this standard could be assessed by the development level of the documents that make up it. It is expected that the standard contemplates all the changes that the control systems have undergone and will undergo in the future.

Hardware root of trust

Different perspectives about the maturity level of hardware root of trust have been found. Shpantzer in 2013 stated that an increasingly complex threat landscape, coupled with the widespread availability and increased maturity of the Trusted Platform Module (TMP) technology, as well as other factors such as cost reduction, makes TPM ready for prime time (Shpantzer, 2013). However, Faisal et al., (2020) affirm that Trusted Computing Group specifications are not mature up till now to secure IoT computing and virtualization.

Digital identity life cycle (asymmetric cryptography)

With regards to digital identity life cycle, the analysis of its maturity level shows some debate. Some opine que the issue of the identity gap is long-lasting since identity systems have evolved over time and needed to mature, while others believe that it is a well-established and developed technology.

3.8.3 EU Project Solutions

Table 43 shows the acronym of the project in which this digital twin solution was developed (the acronym is a link to the project web page), the grant agreement reference (GA), the solution and its description.

Acronym	GA	Solution	Description
certMILS	731456	certMILS develops a security certification methodology for Cyber-physical systems (CPS).	certMILS uses ISO/IEC 15408 and IEC 62443 to develop and applies a compositional security certification methodology to complex composable safety-critical systems operating in constantly evolving hostile environments.
PQCRYPTO	645622	Post-quantum cryptography for long-term security	Design of a portfolio of high-security post-quantum public-key systems, and will improve the speed of these systems, adapting to the different performance challenges of mobile devices, the cloud, and the IoT.
SAFEcrypto	644729	Secure architectures of future emerging cryptography	Practical, robust and physically secure post-quantum cryptographic solutions that ensure long-term security for Information and Communication Technologies (ICT) systems, services and applications. Novel public-key cryptographic schemes (digital signatures, authentication, identity-based encryption (IBE), attribute-based encryption (ABE) will be developed using lattice problems as the source of computational hardness.
CRISP	20847	Cryptographic reduced instruction set processor (CRISP)	New CRISP has been developed which includes a unique security kernel. This guarantees the integrity and data independence of individual applications on a card, delivering a truly secure multisession environment. The project is developing a complete low-cost smart card integrated circuit (IC) combining high speed and Public Key Infrastructure (PKC) security, with the high performance and silicon area efficiency of a reduced instruction set computing (RISC) architecture.
Cathedral	695305	Post-Snowden Circuits and Design Methods for Security	Comprehensive set of circuits and design methods to create next generation electronic circuits with strong built-in trust and security.
SUPERCLOUD	643964	User-centric management of security and dependability in clouds of clouds	New security and dependability infrastructure management paradigms that are: 1) user-centric, for self-service clouds-of-clouds where customers define their own protection requirements and avoid lock-ins; and 2) self-managed, for self-protecting clouds-of-clouds that reduce administration complexity through automation.

Table 43. Security Technologies EU related-results.

3.8.4 Tools

This section is divided into 2 subsections corresponding to (i) digital identity life cycle (asymmetric cryptography) and (ii) hardware root of trust.

The tools for **digital identity life cycle (asymmetric cryptography)** are shown in **Table 44**.

Type	Tools
Protocols using asymmetric key algorithms	• Secure / Multipurpose Internet Mail Extension (S/MIME) • GNU Privacy Guard (GPG), an implementation of Open Pretty Good Privacy (OpenPGP), and an Internet Standard • Europay, Mastercard, and Visa (EMV), EMV Certificate Authority (CA) • IPsec (Internet Protocol security) • PGP (Pretty Good Privacy) • ZRTP (Z and Real-time Transport Protocol), a secure VoIP (Voice over Internet Protocol) protocol • Transport Layer Security standardized by IETF (Internet Engineering Task Force) and its predecessor Secure Socket Layer • SILC (Secure Internet Live Conferencing) • SSH (Secure SHell) • Bitcoin • Off-the-Record Messaging • DSS ("Digital Signature Standard") with the algorithm DSA ("Digital Signature Algorithm") • PGP o Pretty Good Privacy • SSH o Secure Shell
Asymmetric key algorithms and technologies	• Diffie-Hellman • RSA • DSA • ElGamal encryption • Elliptic curve cryptography • Merkle–Hellman knapsack cryptosystem • Goldwasser-Micali • Goldwasser-Micali-Rivest • End-to-end encryption o E2EE
Life cycle management protocols	• CMP (Certificate Manament Protocol • SCEP (Simple Certificate Enrollment Protocol) • ACME (Automated Certificate Manamgement Environment) • EST (Enrollment over Secured Transport).

Table 44. Digital Identity Life Cycle Tools.

The components for hardware root of trust are shown in **Table 45** (Elias, 2021).

Component	Functions
Security perimeter	It defines what needs to be protected on the Security Operations Center (SOC). It can be designed in different ways, for example through a private bus that connects to the main bus by means a gateway.
HSM (Hardware Security Module)	Tamper resistant hardware equipment with security mechanism to store keys and perform cryptographic operations. This equipment is typically certified by a security standard and exposes the operations through a standarized Public-Key Cryptography Standards PCKS11 interface.
Secure central processing unit (CPU)	To run secure software/firmware.

Runtime memory	To protect the runtime data required by the software run in the CPU, specifically the STACK, HEAP and global data.
Tamper resistance	The code that come from the outside needs a validation before running it on the CPU. There are many ways to do so, for example using a dedicated Read Only Memory (ROM) only accessible by the hardware Root of Trust.
Hardware cryptographic accelerators	Although most cryptographic functions are designed using software, cryptography usually uses fewer memory resources and runs faster with the use of hardware accelerators.
True Random Number Generator (TRNG)	To produce a high level of entropy required for the various security functions.
Secure clock, or secure counter	A secure clock is required when reliable time measurement is needed. To do so the hardware Root of Trust should only have access to a clock source that cannot be tampered with.
Secure storage	Secure access to persistent storage is essential for applications requiring a state knowledge. For example, an anti-rollback feature for devices can only truly be secure if the hardware Root of Trust has secure access to a non-volatile memory (NVM).

Table 45. Hardware Root Tools.

3.8.5 Benchmarking and Assessment

This section performs a global assessment of the security technology in general (***Table 46***).

Dimensions	Criteria	Security Tech	Comments
General	Relevance	+++	
	Capability/Features	+++	
	Genericity	+++	
	Performance	++	
	Coverage	+++	
Technological	Interoperability	+	
	Integration	++	
	Scalability	+	
	Security	+++	
	Traceability	+	
Business model	Cost	+	
	Maturity	+	
	Legal Compliance	+++	
	Risk	+++	
	Support	++	
Informational	Need for data traceability	+++	
	Need for data quality	+	
Social	Development-friendliness	-	
	Learning curve	-	
	Social preferences	+	
	Training and documentation	+	

Table 46. Security Technologies Benchmarking.

3.8.6 Application for i4Q Solutions

Table 47 lists the i4Q project solutions that could potentially use security technologies.

i4Q Solution	Description
i4Q^{SH}	i4Q IIoT Security Handler
i4Q^{CSG}	i4Q Cybersecurity Guidelines

Table 47. Security Technologies Application for i4Q Solutions.

4. Benchmarking Analysis among the Digital Technologies

Eight technologies were analysed using six dimensions and several criteria, in order to find which of these technologies meet more criteria and could be more easily used during the development of i4Q solutions. Technologies were thoroughly analysed in Section 3. Here, in Section 4, collected results will be analysed and explained.

For technologies'/tools' and solutions' evaluation, two analyses were conducted; namely, per criteria and per technology. In the next paragraphs both analyses are presented.

Table 48 illustrates the evaluation of the technologies according to the level at which each of the proposed criteria is met, using symbols such as '+++, ++, +, -, --, ---'. For this evaluation, a scale of rating is proposed, in order to quantify the results and have a better understanding on the level each technology meets the analysed criteria. The symbols that are used for the evaluation, e.g., '+++, ++, +, -, --, ---', as well as the expression of each symbol in a score/number are both presented in **Table 49**.

Analysis 1: Dimensions and criteria per technology

	Criteria	Blockchain /Hyperledger	Fog/Edge Computing	Data Analytics	Big Data	Machine Learning	IIoT	Digital Twin	Security Technologies	Total
General	Relevance	+++	+++	+++	+++	+++	+++	++	+++	47/48
	Capability/Features	+++	+++	+++	+++	+++	+++	++	+++	47/48
	Genericity	++	+++	+++	++	+++	++	-	+++	42/48
	Performance	+++	+++	+++	+++	+++	++	+	++	44/48
	Coverage		+++	+++	+++	+++	+++	+++	+++	42/42
Technological	Interoperability	++	+++	+++	+++	+++	++	-	+	35/48
	Integration	+++	++	+++	+++	+++	++	-	++	42/48
	Scalability	++	+++	++	++	++	-	++	+	38/48
	Security	+++	+	++	++	+	-	-	+++	36/48
	Traceability		+	+++		+++	++		+	19/30
Business model	Cost	++	++	++	++	++	++	--	+	36/48
	Maturity	+++	+++	+++	+++	+++	+++	-	+	43/48
	Legal Compliance	+	+	+++	+	+	++	++	+++	38/48
	Risk	+	++	++	++	++	-	+	+++	37/48
	Support	++	+	+++	+++	+++	++	+	++	41/48
Informational	Need for data traceability	+++	+	+++	++	+++	+	+	+++	41/48
	Need for data quality	+++	+	+++	+++	+++	+	-	+	39/48
Social	Development-friendliness	++	+++	+++	++	+++	++	+	-	40/48
	Learning curve	++	--	+	+++	++	++	+	-	34/48
	Social preferences	++	+++	+++	+++	+++	++	-	+	41/48
	Training and documentation	++	--	+++	+++	++	++	++	+	38/48
	Total	**101/114**	**103/126**	**120/126**	**111/120**	**117/126**	**101/126**	**78/120**	**101/126**	

Table 48. Criteria and Technologies; in which level proposed criteria are met by the analysed technologies.

The score that is proposed for the quantification of the data presented in **Table 48** is the following one:

Symbol	Score
---	1
--	2
-	3
+	4
++	5
+++	6

Table 49. Symbols and their score, in order to quantify results.

After the quantification table, two studies were performed:

Study 1. Per criteria.

First study conducted is based on the criteria and the percentage at which these are met by each technology. More specifically, the total score of each one of the criteria analysed in Section 3 of deliverable D1.2, can be used to understand which of these criteria are better fulfilled by the analysed technologies.

If one calculates the final score of each one of the criteria (horizontally/in row calculation), could see which criteria are the most important ones for the technologies analysed. Results show that the best-fulfilled criteria are: relevance and capability from the general criteria, while integration is the best fulfilled from the technological criteria, followed by scalability. Maturity and support are the best fulfilled from the business model criteria, need for data traceability among the informational criteria, followed by the need for data quality and finally, social preferences and development-friendliness are the best-fulfilled among the social criteria.

Relevance and Capability are fulfilled by all technologies analysed at the highest possible level.

It is interesting to notice that some cells in the above-presented table are empty and coloured in grey as well. Criteria in these cells do not apply to the specific technologies (in this case these cells should not be considered for quantification).

Study 2. Per technology.

Another way of analysing the technologies has been performed by calculating the final score of each technology and see which one best meets all the criteria of the bechmarking analysis that was carried out. This can be done by calculating the final score of each technology vertically/by columns.

As **Table 48** shows, Data Analytics, Machine Learning and Big Data are the three technologies that meet most criteria, with a quite good rating. Next, Blockchain / Hyperledger also meets several criteria. Data Analytics, Machine Learning, Fog/Edge Computing, IIoT and Security Technologies fulfill all criteria. Data Analytics and Machine Learning have a positive rating in all criteria, marked by '+++, ++, +' symbols, while for IIoT, Fog/Edge Computing and Security Technologies, it seems that some criteria are not fulfilled in great extent. Digital Twin meets several criteria; however, the final rating of this technology is the lowest among the others.

It is interesting to notice also here that some cells are empty and marked with grey colour. This means that these technologies either do not meet the specific criteria or the criteria do not apply to these technologies.

Analysis 2: Technologies and Tools per i4Q Solutions

Based on the information provided by the i4Q solutions developers regarding the potential technologies that the project solutions will require for their development, it seems that the most demanding ones are data analytics/visualisation and machine learning, since, at this phase of the project, eight different i4Q solutions will take profit of these two technologies. Big data followed by IIoT are the technologies that appear in 3rd and 4th positions, respectively. Therefore, it seems that the BUILD WPs will mainly need for the development of the i4Q solutions, data collection and the extraction of information from data collected, as well as optimisation mechanisms such as statistical algorithms for prediction and classification based on this information. Moreover, based on these results, it is also important to highlight that the i4Q solutions need to manage big amounts of data as well as the appropriate means for collecting such data. For this reason, the results show that the digital technologies related to big data and IIoT are also very relevant. Fog/Edge Computing and Security Technologies are considered also important means for the development of the project solutions, although they will be used in a lesser extent. Finally, it is expected that blockchain/hyperledger and digital twin technologies will be also applied for the development of i4Q solutions, but in this case in the particular development of i4Q^{BC} and i4Q^{DT}, respectively.

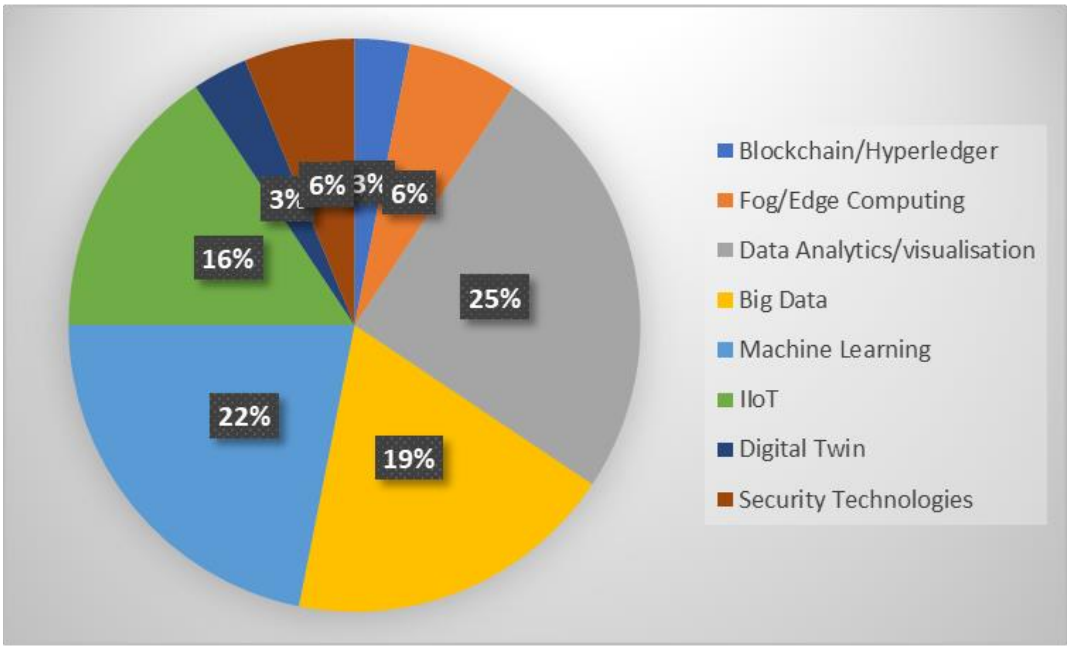

Figure 4. Percentage of the different technologies that will be used for the i4Q solutions development.

Table 50 shows this aggregated analysis in which for each i4Q solutions, the potential digital technologies have been identified and assessed (section 3) and selected (section 5).

#	i4Q Solutions	Technologies							
		Blockchain/Hyperledger	Fog/Edge Computing	Data Analytics/Visualisation	Big Data	Machine Learning	IIoT	Digital Twin	Security Technologies
1	i4Q QualiExplore for Data Quality Factor Knowledge (i4Q^{QE})				Y				
2	i4Q Data Quality Guidelines (i4QD^{QG})								
3	i4Q Blockchain Traceability of Data (i4Q^{BC})	Y					Y		
4	4Q Trusted Networks with Wireless & Wired Industrial Interfaces (i4Q^{TN})						Y		
5	i4Q IIoT Security Handler (i4Q^{SH})								Y
6	i4Q Cybersecurity Guidelines (i4Q^{SG})								Y
7	i4Q Data Repository (i4Q^{DR})						Y		
8	i4Q Guidelines for building Data Repositories for Industry 4.0 (i4Q^{DRG})								
9	i4Q Data Integration and Transformation Services (i4Q^{DIT})			Y		Y			
10	i4Q Services for Data Analytics (i4Q^{DA})				Y	Y			
11	i4Q Big Data Analytics Suite (i4Q^{BDA})				Y				
12	i4Q Analytics Dashboard (i4Q^{AD})			Y					
13	i4Q AI Models Distribution to the Edge (i4Q^{AI})		Y				Y		
14	i4Q Edge Workloads Placement and Deployment (i4Q^{EW})		Y				Y		
15	i4Q Infrastructure Monitoring (i4Q^{IM})			Y	Y	Y			
16	i4Q Digital Twin Simulation Services (i4Q^{DT})			Y		Y		Y	
17	i4Q Data-driven Continuous Process Qualification (i4Q^{PQ})			Y	Y	Y			
18	i4Q Rapid Quality Diagnosis (i4Q^{QD})				Y	Y			
19	i4Q Prescriptive Analysis Tools (i4Q^{PA})			Y		Y			
20	i4Q Manufacturing Line Reconfiguration Toolkit (i4Q^{LRT})			Y		Y			
21	i4Q Manufacturing Line Reconfiguration Guidelines (i4Q^{LRG})			Y					
22	i4Q Manufacturing Line Data Certification Procedure (i4Q^{LCP})								

Table 50. Matching between digital technologies and i4Q solutions (Y: Yes, potentially used for the i4Q development).

Based on this previous aggregated analysis, the study of the tools per technology was performed:

Data Analytics/visualisation

In the block of data analytics/visualisation, seven tools were studied as it is shown in **Table 51**. The potentially most suitable tool for the development of eight of the i4Q solutions is the use of

Python libraries, as it has been chosen for all the i4Q solutions that required this technology for their development.

R libraries, Jupyter and Grafana seem to be also suitable tools for 26,3% of the i4Q solutions. Finally, it is worth mentioning the features of Apache Superset have been considered as potentially appropriate for the development of i4Q^AD^.

	i4Q Solutions							
	9	12	15	16	17	19	20	21
	i4Q^DIT^	i4Q^AD^	i4Q^IM^	i4Q^DT^	i4Q^PQ^	i4Q^PA^	i4Q^LRT^	i4Q^LRG^
Data Analytics/visualisation	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Tableau Software								
Microsoft azure								
Python libraries	X	X	X	X	X	X	X	X
R libraries	X		X	o	X	o		
Grafana	X	X		o		o		
Jupyter	X	X		o		o		X
Apache Superset		X						

Table 51. Matching between data analytics/visualisation tools and i4Q solutions (X: selected; o: optional).

Machine Learning

In the case of the digital technology of machine learning, Python is the most selected tool to develop different i4Q solutions (**Table 52**). Other tools such as R, Apache Spark MLlib, and WEKA are also considered for the project development.

	i4Q Solutions							
	9	10	15	16	17	18	19	20
	i4Q^DIT^	i4Q^DA^	i4Q^IM^	i4Q^DT^	i4Q^PQ^	i4Q^QD^	i4Q^PA^	i4Q^LRT^
Machine Learning	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
Apache Spark MLlib	X	X	X					
R	X			o	X		o	
WEKA		X						
Python	X	X	X	X	X		X	X
Mahout								

Table 52. Matching between machine learning tools and i4Q solutions.

Big Data

The most big data related-tool selected is Tensorflow, as four of the six i4Q solutions that will use this technology has chosen this tool, as it is shown in **Table 53**. Other alternatives considered for the project development involve tools such as keras, Apache Flick and Storm.

	i4Q Solutions					
	1	10	11	15	17	18
	i4Q^QE	i4Q^DA	i4Q^BDA	i4Q^IM	i4Q^PQ	i4Q^QD
Big Data	**X**	**X**	**X**	**X**	**X**	**X**
Microsoft Azure						
Tensorflow	X	X	X		X	
Keras		X			X	
FI-WARE Orion Context Broker						
Apache Flink		X	X			
Apache Storm		X	X			
Apache Spark		X	X			

Table 53. Matching between big data tools and i4Q solutions.

IIoT

From the IIoT technology, only one i4Q solution, specifically **i4Q^TN** will need IIoT hardware devices, as it is shown in **Table 54**. Among these devices different options such as BeagleBoard, Raspberry, OpenMote B, Zolertia RE-Mote have been considered as adequate. With regards to the software and/or operating systems, the most selected one has been Linux. Other alternatives contemplate Contiki, Software Defined Networking Wise (SDNWise) and Open Network Operating System (ONOS).

	i4Q solutions				
	3	4	7	13	14
	i4Q^BC	i4Q^TN	i4Q^DR	i4Q^AI	i4Q^EW
IIoT	**X**	**X**	**X**	**X**	**X**
Hardware		X			
Arduino					
BeagleBoard		X			
Microduino					
Node MCU (ESP 8266)					
OLinuXino					
Raspberry		X			
Tessel					
OpenMote B		X			
Zolertia RE-Mote		X			
Software, operating systems	X	X	X	X	X
Contiki		X			
OpenWrt					
Tiny OS					
Linux	X	X	X	X	X
SDNWise		X			
Open Network Operatin System (ONOS)		X			

Table 54. Matching between IIoT tools and i4Q solutions.

Fog/Edge Computing

From the analysed tools, only IBM Edge Application Manager has been selected as potential tool to develop the i4Q solutions: i4Q^{AI} and i4Q^{EW}.

Security Technologies

In the case of the security technologies, the development of the i4Q solutions: i4Q^{SH} and i4Q^{CSG} will likely be based on 62443 standard, digital identity life cycle (asymmetric cryptography) and hardware root of trust.

Blockchain/Hyperledger

Hyperledger Fabric has been selected as a potential tool due to its functions and features to fulfill the development requirements of the solution i4Q^{BC}.

Digital Twin

In the case of digital twin, different tools have been considered for the development to the solution i4Q^{DT}, such as Ditto Project, PyFMI and/or FMI compliant simulation Tool (i.e., OpenModelica).

5. Application of the Digital Technologies in the i4Q Solutions

5.1 i4Q QualiExplore for Data Quality Factor Knowledge (i4Q^{QE})

QualiExplore is a software service for the i4Q platform that means to increase the users' awareness for information quality. It visualizes information quality characteristics and quality factors relevant for production. The software adopts the Evolutional Data Quality Concept (EDQC) of Liu and Chi (2002). These authors developed a theory-based view on data quality that focuses on the evolution of data along a life cycle. Their data evolution life cycle contains four phases:

- Data collection concerns data capturing through observation of *real-world* processes, measurement, and perception.
- Data organization means structuring and storing of data in files, databases and other forms of data storage.
- Data presentation subsumes processing, interpretation, summarizing, formatting and presentation of data in views.
- Data application is the final phase where users utilize data to achieve a *goal*, which can trigger further data collection.

An important aspect of the EDQC concept is that quality characteristics in a phase contribute to the characteristics of the following phases.

The QualiExplore implementation provides a *2-staged user interface* to support learning about production-related information quality and the factors that influence it. The first stage serves as a filter because the high number of factors can cause information overload for users. Relevant *filter categories* are the user's goals, quality (information characteristics), and channels/sources. The goals include the perspective of the information user and the information creator/author. This is useful because it emphasises that many measures to avoid quality problems require the involvement of both parties. Each category has several statements that represent areas where the user should be or might want to be aware about information quality problems and its related factors. The indicated *factor categories* structure the factors and provide a link between statements and factors.

In i4Q, QualiExplore will use a (graph) database in the backend. Its frontend will consist of a visual interface to show and edit factor knowledge (i.e., static descriptions of data quality factors organized in a tree-structure), and a conversational interface (i.e., chatbot) to simplify and speed-up the access to factor knowledge. The solution currently uses a MongoDB as database, but BIBA plans to use a graph database (e.g., Neo4j). The chatbot bases on the Rasa technology stack (open source) and uses Tensorflow.

5.2 i4Q Data Quality Guidelines (i4Q^{DQG})

This is a text document and does not require any technology support.

5.3 i4Q Blockchain Traceability of Data (i4Q^{BC})

Blockchain technology enables data traceability, which is a capability provided by i4Q^{BC}. Blockchain provides the underlying technology for ensuring immutability, finality, consensus, and provenance contributing to non-repudiation of the stored state. i4Q^{BC} shall provide smart contracts on top of the blockchain infrastructure to provide these capabilities in a smart manufacturing environment, for storing and querying manufacturing related data. This service shall take as a starting point the capabilities of Hyperledger Fabric (https://www.hyperledger.org/use/fabric), which is an infrastructure for permissioned blockchains, exhibiting scalability without scarifying privacy.

5.4 i4Q Trusted Networks with Wireless & Wired Industrial Interfaces (i4Q^{TN})

The i4Q^{TN} solution is one of the base elements of the proposed architecture in i4Q project, since it will guarantee that digitized processes and higher-level platforms can exchange information with a guarantee of reliability and quality of service. The main objectives of the i4Q^{TN} solution are to establish a set of communication technologies focused on two application areas. The first one is aimed at the exchange of large volumes of information wirelessly, with high reliability, high data rates and low latency. The second is focused on the deployment of networks with a multitude of devices at a low cost, implemented through Industrial IoT systems. The main requirements for the i4Q^{TN} solution are exposed below:

- Select a communication infrastructure that improves the reliability and quality of the data exchanged between the plant processes.
- Use mechanisms and technologies that allow wireless networks to achieve latencies similar to those obtained in wired networks.
- Complement the aforementioned communication infrastructure with IIoT technologies, using industry-oriented WPAN (Wireless Personal Area Network) technologies, which allow increasing the digitization points (high scalability) at a low cost.
- Integrate the SDN paradigm in WPAN networks to improve the quality of service, establishing guaranteed information flows in mesh networks, as well as allowing node mobility while reducing management complexity.

Regarding the technologies and tools to be used, the i4Q^{TN} solution focuses on two different communication technologies. On the one hand, the use of low-power wireless technologies such as WPANs, facilitates digitization processes through a large number of devices at a low cost. In this sense, the Contiki-NG operating system will be used for the development of a reliable protocol stack, focused on industry thanks to the TSCH medium access method. Along with this system, the software SDN-WISE (Software Defined Networking solution for Wireless Sensor Networks) will be implemented, in order to simplify the management and scheduling of the WPAN radio resources, combined with the ONOS (Open Network Operating System) controller, which allows a fully centralized resource management.

On the other hand, it is proposed to assess and eventually deploy a private 5G network infrastructure, which considerably improves reliability and quality of services, compared to fourth generation technologies. In addition, thanks to the use of private infrastructure, high availability and scalability is achieved, as well as increasing the network traffic capacity and the efficiency of the network. It is also proposed to complement this 5G private networks with TSN (Time-Sensitive

Networking) mechanisms, in such a way that these wireless communication over 5G networks achieve latencies similar to those presented by wired communication technologies.

5.5 i4Q IIoT Security Handler (i4Q^SH)

i4Q^SH is a cloud service that distributes trust across the architecture using a hardware secure module as a trust anchor point. Once the trust is distributed, the software enables the mechanisms to expose cryptography operations that other i4Q Solutions can consume, adjusting security and safety policies at different levels to ensure trustability and privacy of data.

i4Q^SH cloud service offers two different endpoints:

- CA with HSM

CA with HSM endpoint will provide a Certificate Management Protocol (CMP) interface to manage the life cycle of digital identities of different entities.

CMP is a well-known protocol defined in Request for Comments (RFC) 4210 and RFC 6712 when the transport of the protocol is Hypertext Transfer Protocol (HTTP).

The image **Figure 5** from the RFC 4210 defines the interaction points of the CMP protocol when the digital identity life cycle is managed.

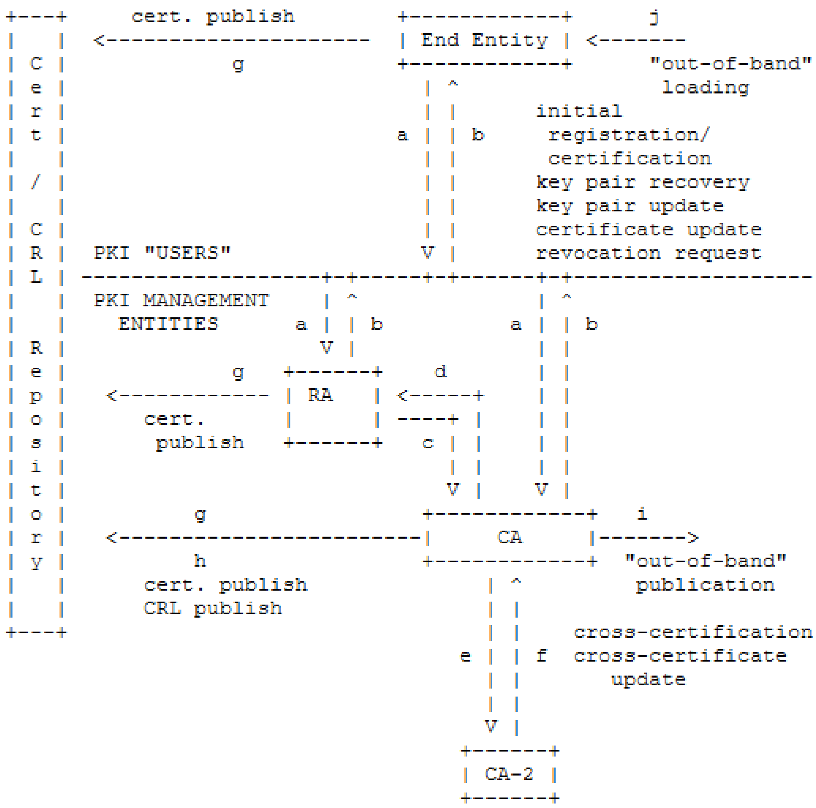

Figure 5. Operations managed by CMP during the digital identity life cycle.

- PKCS11 Rest Service with HSM backend

PKCS11 interface defines the following functions:

Characteristics	Description
General purpose	C_Initialize(), C_Finalize(), C_GetInfo(), C_GetFunctionList()
Session management	C_OpenSession(), C_CloseSession(), C_GetSessionInfo(), C_CloseAllSessions(), C_Login(), C_Logout()
Slot and token management	C_GetSlotList(), C_GetSlotInfo(), C_GetMechanismList(), C_GetMechanismInfo(), C_SetPIN()
Encryption and decryption	C_EncryptInit(), C_Encrypt(), C_EncryptUpdate(), C_EncryptFinal(), C_DecryptInit(), C_Decrypt(), C_DecryptUpdate(), C_DecryptFinal()
Message digesting	C_DigestInit(), C_Digest(), C_DigestKey(), C_DigestUpdate(), C_DigestFinal()
Signing and applying MAC	C_Sign(), C_SignInit(), C_SignUpdate(), C_SignFinal(), C_SignRecoverInit(), C_SignRecover()
Signature verification	C_Verify(), C_VerifyInit(), C_VerifyUpdate(), C_VerifyFinal(), C_VerifyRecoverInit(), C_VerifyRecover()
Dual-purpose cryptographic functions	C_DigestEncryptUpdate(), C_DecryptDigestUpdate(), C_SignEncryptUpdate(), C_DecryptVerifyUpdate()
Random number generation	C_SeedRandom(), C_GenerateRandom()
Object management	C_CreateObject(), C_DestroyObject(), C_CopyObject(), C_FindObjects(), C_FindObjectsInit(), C_FindObjectsFinal(), C_GetAttributeValue(), C_SetAttributeValue()
Key management	C_GenerateKey(), C_GenerateKeyPair(), C_DeriveKey()

Table 55. PKCS11 interface functions

The REST API endpoint will expose the cryptography functions based on the identity of the entity that requests the REST operations.

5.6 i4Q Cybersecurity Guidelines (i4Q^SG^)

i4Q^CSG^ is a document that follows 62443 standard and defines a set of recommendations to enable multilayer cyber security features in IIoT, enabling IIoT devices to interact with industrial platforms securely in all stages in a manufacturing scenario.

i4Q^CSG^ also includes an architecture and methodology to provision signed certificates with Hardware Security Module (HSM) and trusted material to devices' Trusted Platform Module (TPM).

As a document, this solution does not require any technology support.

- IEC 62443 - Security for Industrial Automation and Control Systems

IEC 62443 is a set of documents covering the cybersecurity needs for a complete solution in the field of industrial control and automation systems.

The figure below(extracted from IEC 62443-4-1) represents the scope of influence of each of the documents within a complete solution.

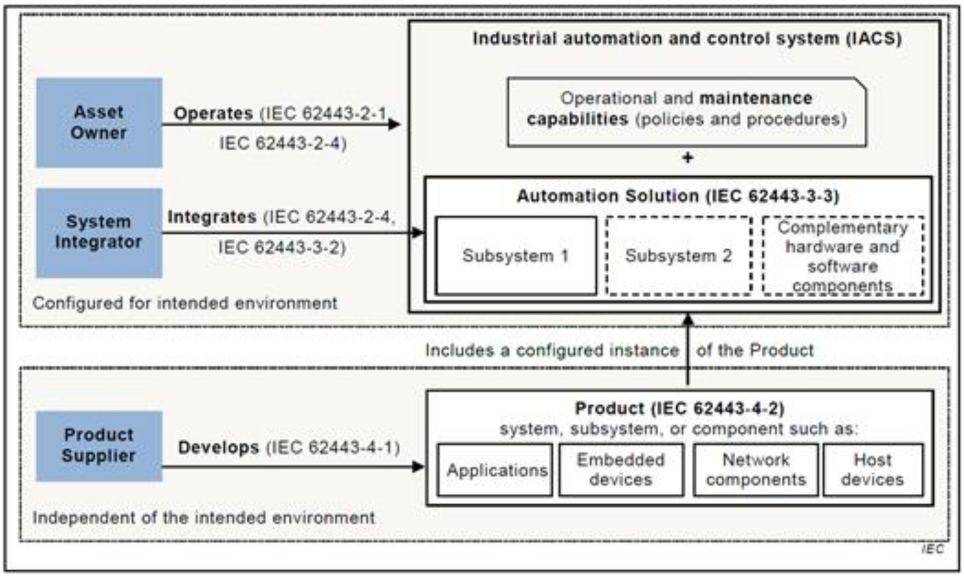

Figure 6. Example of the scope of the IEC 62443 standard documents.

The figure shows three different roles: product supplier, system integrator and asset owner. Each of them must follow certain parts of the standard to perform their role with all cybersecurity considerations.

A product supplier is the one who provides some component to be integrated in the solution. In the case of the wind turbine, it can be the drive control unit, a controller (PLC), a Human-Machine Interface (HMI), or a network device (switch, firewall). Some of these components may be from third parties, while others may be developed by the integrator itself.

- TPM and Hardware Security Module (HSM) as hardware root of trust to provide trusted digital identities

Digital identities provide the trust when communicating different entities or when data travels through different domains between different entities. The procedure to provide a digital identity in a secure way may ensure that the identity cannot be spoofed, is unique and is for the entity the identity belongs to. This could be achieved by using secure hardware among the process, from the first issue of the identity until it is correctly installed in the corresponding entity. Asymmetric cryptography and Public Key Infrastructure (PKI) help to distribute digital identities between different entities. The Public Key Infrastructure may have a hardware secure module to store the private part of the digital entities that are used to verify and validate the end entity identity when a new certificate is issued. At the same time, the end entity may store the private part of digital identity in a trusted platform module (TPM) to assure that the private part is secret and immutable.

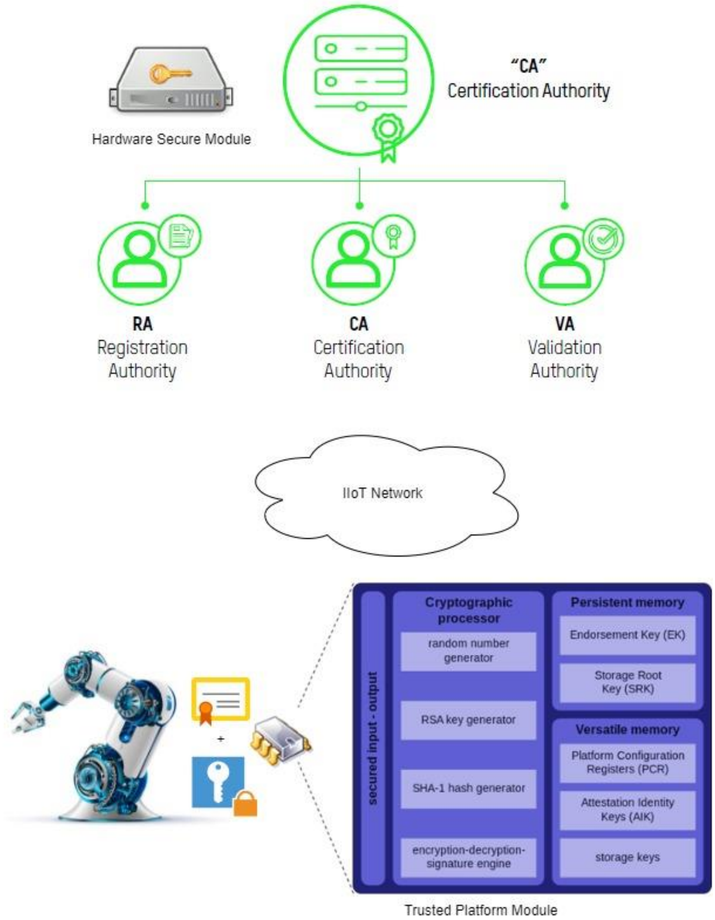

Figure 7. Example of HSM within PKI and TPM within IIoT device

The figure above shows how the HSM is within the PKI and the TPM within the IIoT device to provide hardware root of trust to the digital identity of the IIoT device.

5.7 i4Q Data Repository (i4Q^DR^)

The data repository will be the service in charge of receiving, storing and serving the data to the other components in the architecture in a proper way. In the grand scheme of the system, we can understand the data repository as the place where all the collected data will reside, ready to be consulted or consumed.

In order to specify the relevant technologies for i4Q^{DR} we must establish the functional boundaries of this particular task. i4Q^{DR} focuses on the data repository itself. Some other aspects will need to be taken into account, as the communication channels for the repository or the way information will be stored. However, there are other tasks that will handle those specific problems, such as i4Q^{DIT} or i4Q^{TN}. These tasks are closely related to i4Q^{DR} and its components will interact with it, but the specific relevant technologies for them will be discussed on their respective sections.

Once the scope of the solution is defined, the use case needs to be addressed. The most relevant technologies will be the ones that better fit the desired use case. There are no technical estimations or requirements describing the load and size of the final system. Once this information is known, better approaches can be taken to select the best fitting technologies. It can be assumed that the repository will be in charge of receiving large amounts of data from multiple sources, thus, requiring both a high throughput and a great amount of space. In order to handle such demands, as well as its variations, the system must be provisioned adequately. The most sensible way to do so is through elasticity. An elastic repository can address the needed storage or replication changes and adapt to them automatically, being able to scale in and out as required. In order to achieve this elasticity, the different layers of the solution will be considered:

- Hardware layer

The way the hardware will be set up is a relevant factor in order to achieve a good solution. If the hardware is physically distributed, all the centralised approaches are ruled out. On the other hand, if a central system is considered, distributed technologies can still be used successfully.

- OS Layer

The system can be partitioned at an OS-Level through virtualisation. This approach is highly recommended, as it will help to better assign the available computing resources. The system can achieve elasticity through this type of virtualisation. However, there is a big overhead associated to the creation of a virtual machine, which is not optimal for fast-paced scenarios. Some technologies to consider for the virtualisation will be VirtualBox, Openstack or VMWare. As for the OS itself, any Linux-based server system is suitable, mainly because its compatibility.

- Process Layer

Another way to approach the system partitioning is at a process level, commonly addressed as containerisation. This approach is much more suitable for a rapid-changing environment, as containers can be created and destroyed swiftly. An orchestration mechanism will need to be used in order to control the container behaviour. Some relevant technologies on this matter can be found in D1.1.

- Public Cloud Layer

A Hybrid cloud scheme might be relevant to consider if the infrastructure on premises is not able to handle all the incoming traffic due to load variations or surges or even if the volume of data in the repository grows unexpectedly fast. The system will be able to off-source all the computing or storing load that exceeds its current power to the public cloud infrastructure. The integration of one of these public cloud providers on the system is highly dependent on the chosen virtualisation, and it will consist of replicating the building blocks on premise (VMs or containers) on such infrastructure. If the storage needs of the repository grow constantly over time, the Hybrid cloud scheme stops making sense, as its instances are thought to be temporal. An expansion of the on-premises infrastructure will be more suitable on that case, both from a performance and economic standpoint.

As for the application itself, some technologies can be considered depending on the desired architecture. If the load is big enough to justify a distributed system, any of the most common distributed database management systems are relevant. Some examples can be Mongo Data Base (MongoDB), Cassandra, Oracle or CockroachDB. The selection of one of them will mainly depend on the data model, as the way the information is stored will affect its performance. On the other hand, if the data income can be supported by a centralised system and the architecture allows it, other solutions as PostgreSQL (Postgre Structured Query Language) or MySQL can be considered. Other technologies can be useful for the communications, such as REST APIs or Messaging interfaces. Their use will be related to the data gathering mechanisms and the information flow among the system. Some technologies relevant in these categories are Express or Flask (Rest APIs) and Kafka, RabbitMQ or ZMQ (messaging).

5.8 i4Q Guidelines for building Data Repositories for Industry 4.0 (i4Q^DRG)

The i4Q^DRG task consists of the composition of text document. It will not require any specific technology to be completed.

5.9 i4Q Data Integration and Transformation Services (i4Q^DIT)

The i4Q^DIT will be one of the base solutions, quite probably supporting all pilot cases and will interact with most of the other solutions. The main function of this solution will be to provide integrated datasets with homogeneous structure of different recordings (measurements) from multiple sources, so as to prepare the manufacturing data for further analysis. For example, Computer Numerical Control (CNC) data and accelerometer recordings will need to be combined for further analysis. Other functions offered by i4Q^DIT will be:

1. Read different types of data produced by the manufacturing workflows
2. analytic and decision-making services,
3. data preparation: cleaning, preprocessing, feature extraction, data filtering
4. Integration of input data, e.g., early fusion
5. Combination of analysis' results, e.g., late fusion

Regarding the technologies described in Section 3, the following will be used: **data analytics** techniques, for the data preprocessing stage that will include filtering, cleaning, missing values imputation, alignment of time series, alignment of variables that will be used as a key for merging the input of different sources and feature extraction. Apache Storm will be quite useful for the processing of large volumes of streaming sensor data. For the higher-level processes of filtering and feature extraction, Python and R libraries will be used. Grafana may be adopted for visualization purposes. **Machine and deep learning** methods are also crucial for this solution, for example ensemble learning methods may be used for fusion of classification results of different models. Python and R libraries are the tools considered, as well as Apache MLib. . Since the manufacturing workflows produce large amounts of data, **big data** tools will be employed for the deployment of this solution, like Keras.

The data integration solution will be developed, based on other existing solutions, that will be parameterized to fit the project's specific needs. i4Q^DIT intends to use Dask platform, that uses existing Python APIs and data structures. Other existing solutions considered are Talend platform and Spark that can handle large amounts of data.

5.10 i4Q Services for Data Analytics (i4Q^{DA})

The services for data analytics, are responsible for knowledge discovery and data mining tasks. This service layer, working together with processing engines, has the key function of creating value from the stored data by retrieving and showing useful information. This process is able to query, aggregate, correlate, learn and apply cognitive technologies and thereafter show valuable information in a meaningful way. Such information can have a critical impact in production by allowing facilities to improve their global efficiency when manufacturing, through the use of applications and mechanisms that consume this information and adapt the production on the fly. On the other hand, it can contribute for a better overall knowledge of production´s performance and weaknesses, either for real-time assessment or later analysis. Apache Spark MLlib and Grafana are some of the technologies used.

The main core services that could be applied to i4Q are related with:

- Outlier and anomaly detection in quality data

We use two different approaches for anomaly detection: (i) The most basic method to identify anomalies is to use common statistical properties of distribution such as mean, median, mode, and quantiles. This method can be useful for simple use cases where the boundary between normal and abnormal behavior of the process is clearly understood and a single source of failure can be identified, and (ii) The Machine Learning method, which includes the Density Based Anomaly Detection and the Clustering Base Anomaly Detection. Anomaly detection services will enable to answer the following questions: Where and when abnormal behavior has occurred? Where and when an anomaly has occurred or will occur? When maintenance or new equipment is necessary?

- Forecasting services in quality data

Smart manufacturing requires real-time prediction capabilities for use cases such as inventory stocking for just-in-time production, quality monitoring, and anomaly detection. Forecasting or predictive analytics in quality data, propose a classification of these applications, identify the gaps, and provide insights for future research. The approaches used for vary in terms of the nature of the data, but also, in terms of the context and the problem with intend to solve, therefore our suit apply different algorithms such as: time-series forecasting, clustering, K-nearest-neighbors, neural networks, regression analysis, support vector machines, and support vector regression.

- Correlation analysis in quality data

In Zero-Defect manufacturing strategies for production, systems require the knowledge of existing complex correlations from different parameters. Thus, it is possible to derive cause and effect relationships independently of the level of measurement of the gathered data in highly connected processes. Correlation analysis enables to explain how one or more variables in a dataset are related to each other. These variables can be input data features which have been used to forecast our target variable. The correlation analysis services provide different correlation measures such as Pearson, Spearman and Kendall correlation measures.

5.11 i4Q Big Data Analytics Suite (i4Q^BDA^)

Figure 8 shows the general architecture and core technologies that support the big data analytics suite. Namely to efficiently gather, harmonize, store and apply analytic technics to data generated by sensors and other Cyber Physical Systems mounted on facilities. 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.

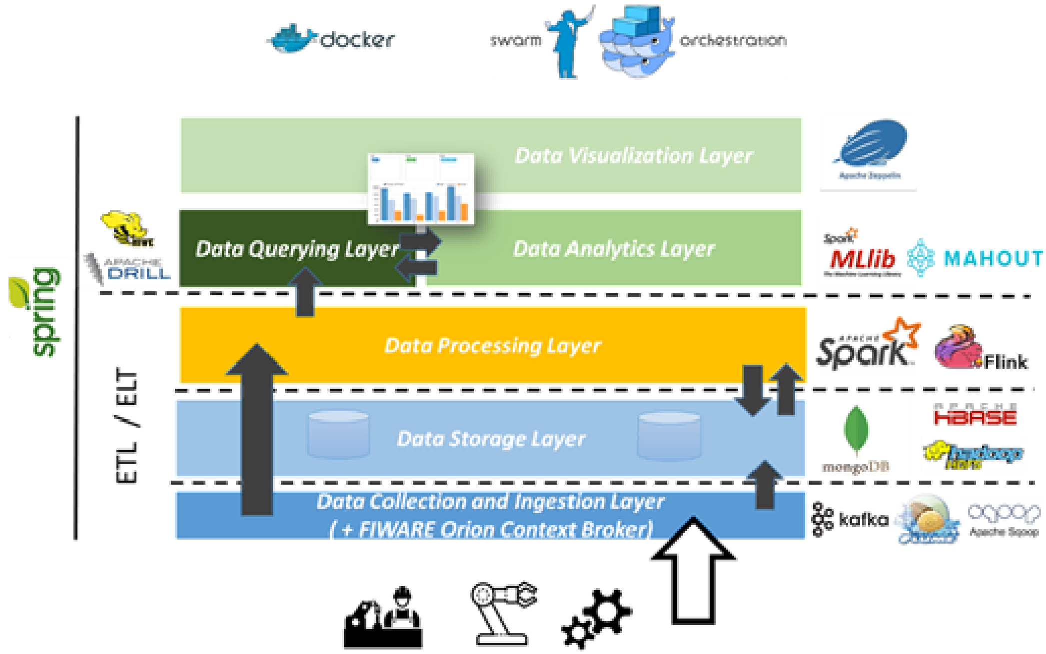

Figure 8. Big data analytics suite architecture.

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. 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 big data suite, 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.

The technologies shown in the above infrastructure are deployed as Docker containers on a distributed Kubernetes environment. As described in D1.1, Docker is a well-known operating system-level virtualization where software packages called containers operate isolated from each other with their own libraries and configurations, communicating with each other through well-defined channels. Each container is a runnable image that include everything needed to run the application, libraries, environmental variables, configuration files, etc. Docker eases continuous integration, reproducibility and isolation, while providing better environment management and scalability. Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications. It groups containers that make up an application into logical units for easy management and discovery.

Docker Swarm tool is used to scale processing/storage capabilities and better benefit from distributed technologies such as Hadoop, Apache Spark, Apache Flink, MongoDB, and others included in the general architecture. Docker Swarm is a clustering and scheduling tool for Docker containers, integrating orchestration capabilities with decentralized design, multi-host networking, Transport Layer Security (TLS) mutual authentication and encryption between all other nodes to secure communications, Multi-host networking and service discovering. Working as a declarative-service model to deploy services to any applications stack. It allows the manage and establishment of a cluster of Docker nodes as a single virtual system, allowing easy implementation of strategies for load balancing, high availability, and perhaps the most advantageous feature, self-adaptation and automatic scalability of distributed technologies to available nodes. That is, giving the example of how Apache Spark works, with a Master and one or more Workers, if a new node joins the Swarm, Docker Swarm can automatically create a new Spark Worker service in the new node and make it available to the Spark Master. Swarm uses Raft Consensus Algorithm to manage cluster state and nodes membership. It fundamentally assures that the state of the cluster is the same in Manager nodes through a node agreement logic, so that tasks and services can be picked up as they were before in case of failure. In such cases, Manager can schedule and re-balance tasks to ensure state recovery. Swarm architecture abstraction is shown in **Figure 9**.

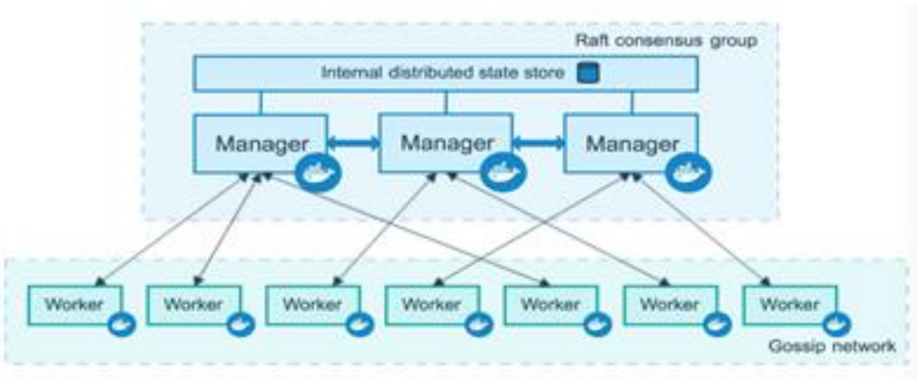

Figure 9. Docker Swarm architecture

Figure 10 shows, for exemplification, how technologies can be distributed across Swarm nodes in Docker containers. Technologies like Hadoop can be deployed to the Swarm as services with a specific description on how and where to run. For instance, Hadoop can have one Data Node Service in each node or just in some, being this distribution automatically adapted if a new node entry or leaves the swarm. Each node can have N containers, running N technologies, working all together.

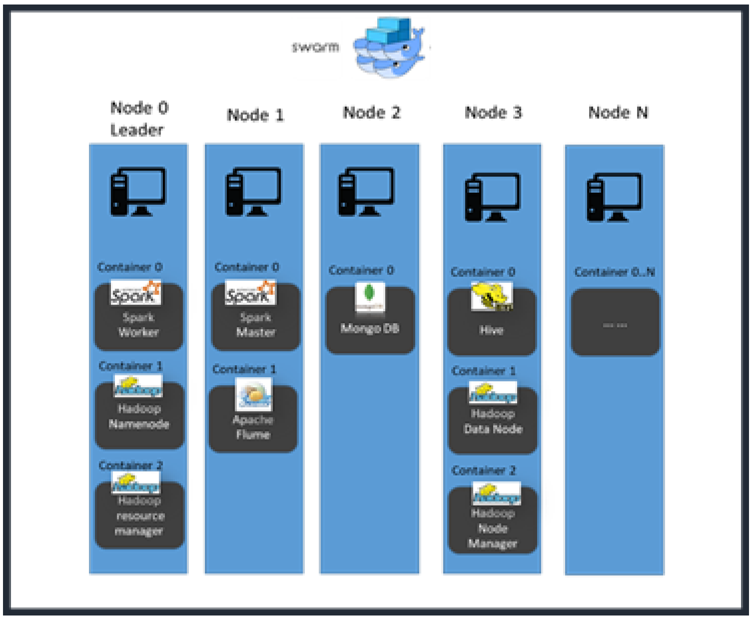

Figure 10. Example of technologies running in several nodes in a Docker Swarm Orchestration

5.12 i4Q Analytics Dashboard (i4Q^AD)

A dashboard is a real-time visual representation, in this case, of manufacturing processes. It displays, typically in graphical or chart form, the key KPI or metrics that indicate performance. Dashboards present data and insights from machines, sensors, and systems in a user-friendly way. They help manufacturers monitor and optimize production quality and efficiency. At the same time, manufacturers gain insight in specific operations. Through dashboards, organizations can quickly see their current and historical performance.

In i4Q we believe that the adoption of intuitive dashboards applied to the manufacturing domain will bring the following value added:

- Make complex information simple
- Give the right information to the right person at the right time
- Gain complete visibility of production lines

The picture below depicts the visualizations of the analytics dashboard, deployed under the big data analytics suite (i4Q^BDA). The presented dashboard was developed in Grafana technology.

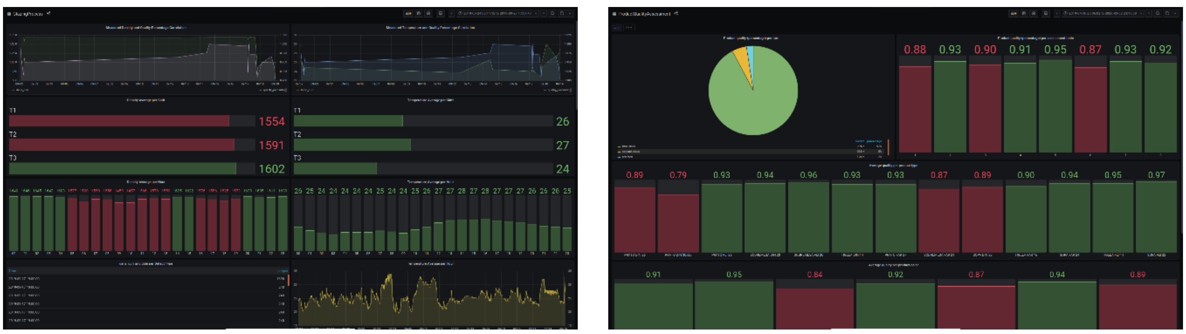

Figure 11. Visualizations of the analytics dashboard, developed in Grafana technology

5.13 i4Q AI Models Distribution to the Edge (i4Q^AI)

Edge technology, as described in 3.2, enables several desirable characteristics for a smarter factory scenario, such as low latency, enhanced security and privacy, reduced network bandwidth utilization, and a fist line of defence in a big data architecture (data reduction, by keeping data local and not necessarily transmitting all data to the cloud). i4Q^AI shall provide such capabilities by distributing AI models to edge nodes, enabling processing at the edge, taking advantage of the associated low latency for enabling faster reaction to data in real-time **Figure 13**.

AI is used for identifying anomalies in a production environment and proposing actions to remedy the situation. An integrated overall AI model life cycle management to run adequately in an edge environment shall be pursued by i4Q^AI. Edge AI based workloads require access to their AI models, which is challenging to keep up-to date when taking into consideration a scalable environment with potentially constrained resources. Edge resident AI models will enable processing of data close to the source in which it was produced, thus opening the door for taking actions faster.

5.14 i4Q Edge Workloads Placement and Deployment (i4Q^EW)

Edge computing, as described in 3.2, is an enabler for addressing performance, latency and security/privacy requirements. Knowledgeable deployment and execution of AI workloads on an edge computing environment is required to efficiently operate on the edge (see ***Figure 13***). Proper placement and deployment services are required to take advantage of the edge environment. A hierarchical Cloud/edge architecture provides efficient and flexible management of edge workloads, to gain from the advantages of the edge environment.

AI: i4Q^EW provides interfaces and capabilities to run different AI workloads on different devices; primarily edge devices. For AI at the edge automated workload management is essential to address scale and dynamic heterogeneity of workloads. Placement, deployment, execution and monitoring cycles shall optimize AI workloads operation on the edge.

5.15 i4Q Infrastructure Monitoring (i4Q^IM)

This solution provides an ensemble of monitoring tools for smart manufacturing workload orchestration and predictive failure alerting, including monitoring the health of workloads and productively alerting and taking corrective actions when a predicted problem is detected.

The technologies of Section 3 that will be used in infrastructure monitoring, are **big data** analytics and **machine/deep learning** methods, for training models on historical data that will later be used to predict failures. The Grafana analytics platform is considered for this solution, since it provides visualization and alerting functions.

5.16 i4Q Digital Twin Simulation Services (i4Q^DT)

i4Q^DT solution allows industrial companies to achieve a connected 3D production simulation, with a digital twin for manufacturing enabling virtual validation/visualisation and productivity optimisation using pre-existing and data from different factory levels (small cell to entire factory).

Regarding **Digital Twin** technologies the i4Q[DT] solution will make use of the Ditto Project, to have a virtual representation and contextualization of all the assets present in the manufacturing line. Thanks to Ditto, the virtual representations of the different physical devices are accessible through APIs, creating a framework of consistent interoperability that allows the building of the digital twins in a more focused digital twin domain, and reducing the complexity of IoT deployments.

Additionally, when virtual sensors are to be obtained, physics-based models will be developed. As these models can be developed using different modelling tools and SW, an industrial model exchange standard is proposed: the FMI. The FMI facilitates the integration of models with monitoring algorithms, protecting the model intellectual property, and making the framework independent from the modelling source.

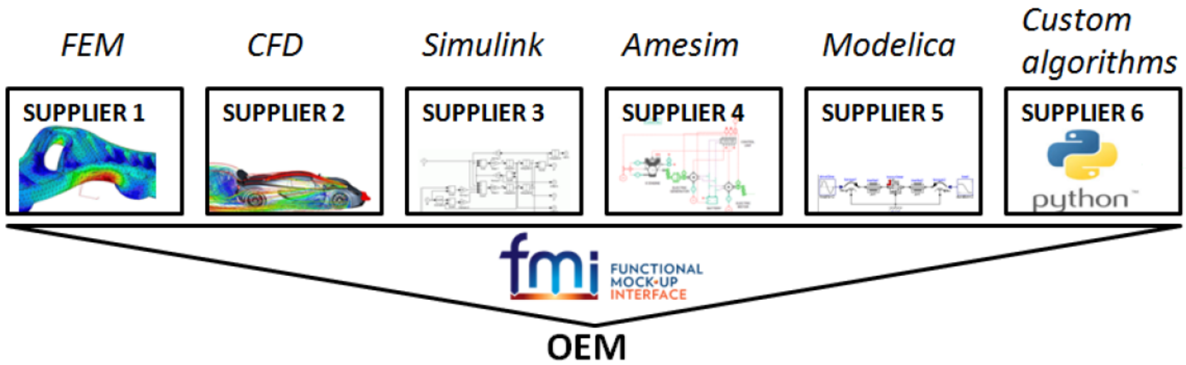

Figure 12. Example of use of the FMI standard (Gonzalez, 2020).

Exporting a model using this standard implies the generation of an FMU, which is a compressed file that contains the definition of the variables and parameters of the model, as well as a set of C-functions. The FMU files can be used to run simulations with two different approaches: model-exchange and co-simulation. In the first scenario, the FMU results in an input/output block that may be used by other modelling and simulation environments; in contrast, in the second scenario, each model is exported with its own solver and it communicates with other sub-models at certain discrete communication points. To ensure synchronization, a master model controls the data exchange.

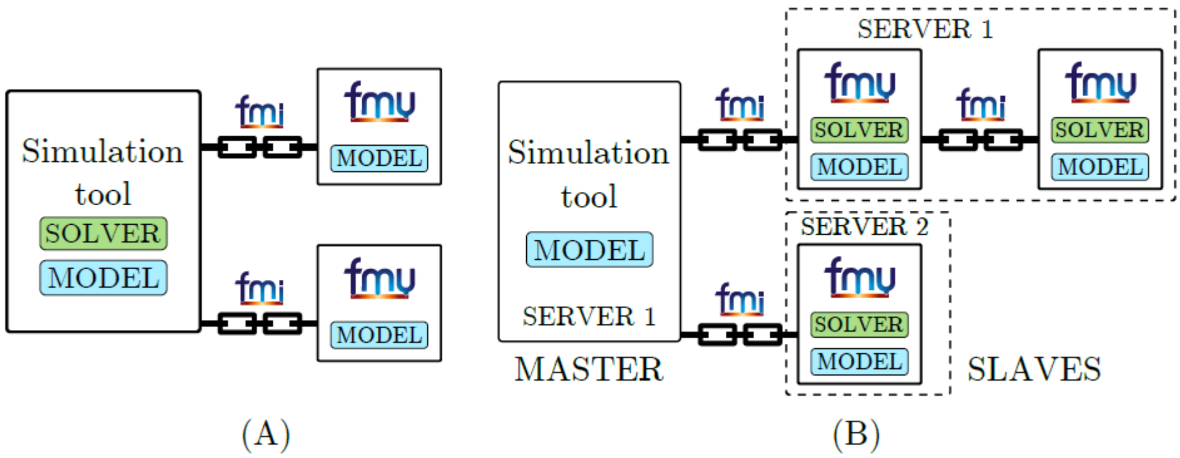

Figure 13. Approaches for running simulations: (A) model-exchange, (B) co-simulation (Gonzalez, 2020).

With the aim of managing the development of the models and the monitoring algorithms, FMI compliant open-source simulation tools like OpenModelica and general-purpose programming languages like Python (through the PyFMI library) can be used.

To virtually replicate the various stages of the manufacturing procedure (both at process and product level) a set of several **data analytics** and **machine learning** functions will be developed, using R and/or Python libraries. Both these platforms offer a broad collection of techniques that allow for supervised and unsupervised classification (regressions, neural networks, support vector machines, clustering methods, etc) as well as time series analysis (estimate uncertain events of the past, predict future events) that help in the development of the mathematical methods that will ultimately be the building blocks for the physical models of the digital twins. Furthermore, being continuously feed with new algorithms developed by the community, both of them allow for an access to state-of-the-art methods in the field of data analysis.

Figure 14. Python libraries for Machine Learning and Data Analytics[8]

In order to make sense of the huge amount of data handled by the digital twin, different **data visualization** tools like Grafana are foreseen. This solution is appropriate for running data analytics, pulling up metrics and monitor them with the aid of different sets of customizable dashboards. Being open-source, Grafana also allows the development of modules for integration with several different data sources, giving extra flexibility in the interaction between the digital models and the real data from the manufacturing line.

Also, some general Python or R options, such as Tkinter or Shiny, are good candidates for data visualization frameworks, as they allow a direct integration with the data analytics functions of each platform. Since 3D representations of the complete production line are also considered, tools like Godot or Unity could also be useful.

5.17 i4Q Data-driven Continuous Process Qualification (i4Q^PQ)

Blockchain technology enables traceability of product data and production data, which is needed for i4Q^PQ to map part quality to process parameters. This is used for process approval via *Pre-*Control for ramp-up procedures to verify if the process is stable and predictable. Further the

[8]https://towardsdatascience.com/best-python-libraries-for-machine-learning-and-deep-learning-b0bd40c7e8c

traceability is used in monitoring of flow process manufacturing. The data is used for continuous and predictive process control.

Machine learning and **data analytics** are methods for developing algorithms that analyse data and find interactions between different characteristics. This is essential for i4Q^PQ because knowledge of process data patterns enables prediction of product and process quality and early detection of deviations. If anomalies are detected by developed algorithms in data provided from the manufacturing process, prescriptive measures are initiated to ensure stable processes and high-quality products. Furthermore, pattern detection is used for predictive maintenance and other cases.

IIoT will provide information and data platform for data exchange and as a data source for i4Q^PQ algorithms and analysing tools. **Big Data** should be provided here in the needed quality as basis for algorithm development and verification.

For i4Q^PQ, a **Digital Twin** of flow production processes is efficient for generating data and testing algorithms without the need to stop production. The Digital Twin technology is time and cost efficient for algorithm development. Algorithms are developed, adjusted and validated on the Digital Twin platform. When the algorithm has reached a high level of maturity it is implemented and adapted to the actual process. Further adjustments are necessary afterwards due to the differences between the digital twin and the real process.

5.18 i4Q Rapid Quality Diagnosis (i4Q^QD)

i4Q^QD is a micro-service for providing rapid diagnosis of manufacturing line on the possible cause of failures, evaluating data fidelity, product-quality and process condition, and providing action recommendations for sensor/data processing recalibrations, process line/machine reconfiguration or maintenance actions.

Rapid quality diagnosis requires image processing and computer vision methods, thus **deep learning methods** and respective tools will be used extensively. **Big data** techniques are also important, to manage the large workflows.

5.19 i4Q Prescriptive Analysis Tools (i4Q^PA)

i4Q^PA solution is a micro-service consisting of simulation models as a service, taking as input the manufacturing resources, current production planning and process condition, and proposing process configuration parameters, ensuring that non-simulation experts may also exploit the prescriptive analyses.

The i4Q^PA solution will make use of different **data analytics** platforms, thus allowing the evaluation and comparison of the results that come from the simulation of different industrial scenarios in the i4Q^DT. For data analytics Python and/or R libraries will be used, as they provide an extensive set of optimisation techniques that help in the development of the mathematical methods that will make the prescriptive analysis. These optimisation techniques will have as objective the selection of the best of the digital twin instances that have been simulated, or even the selection of those simulation scenarios that need to be selected or avoided in order to reach faster the optimal solution, based on a scenarios map. Furthermore, being continuously fed with new algorithms developed by the community, both allow for an access to state-of-the-art methods in the field of data analysis. Previous experiences in developments with R show that the *metaheur* package provides easy-to-implement metaheuristic optimization functions, while the *ga* package contains flexible general-purpose functions to use genetic algorithms. Regarding Python, packages like *scipy* provide the necessary functions for optimization problems.

In order to make sense of the huge amount of data handled by the multiple scenarios, different **data visualization** tools like Grafana are foreseen. This solution is appropriate for running data analytics, pulling up metrics and monitor them with the aid of different sets of customizable dashboards. Being open-source, Grafana also allows the development of modules for integration with several different data sources, giving extra flexibility in the interaction between different instances of the digital twin. Also, some general Python or R options, such as Tkinter or Shiny, are good candidates for data visualization frameworks, as they allow a direct integration with the data analytics functions of each platform.

Inside this data analysis and comparison between scenarios some **machine learning** functions could be developed, also using R and/or Python libraries. Both these platforms offer a broad collection of techniques that allow for supervised and unsupervised classification (regressions, neural networks, support vector machines, clustering methods, etc) that help in the development of the algorithms.

5.20 i4Q Manufacturing Line Reconfiguration Toolkit (i4Q^{LRT})

i4Q^{LRT} solution is a tool that aims to increase productivity and reduce the manufacturing line reconfiguration effort through AI. This toolkit consists of a set of analytical components to solve known optimisation problems in the manufacturing process quality domain by finding the optimal configuration for the modules and parameters of the manufacturing line. It will allow to set up machines along the line focusing on quality standards, using as a basis some examples of problems that i4Q^{LRT} solves for manufacturing companies.

To improve productivity, i4Q^{LRT} will use different Python libraries that enables **data analytics** that can be obtained directly from the manufacturing line. Some optimisation libraries that are used are SciPy (Non-Linear Optimisation), networkx (graph algorithms), OR Tools (Routing, Continuous Linear Programming, and Integer Linear Programming). Thanks to this data, by means of **machine learning** tools such as Pytorch or Keras, i4Q^{LRT} will offer different reconfiguration solutions for the production line.

5.21 i4Q Manufacturing Line Reconfiguration Guidelines (i4Q^{LRG})

i4Q^{LRG} solution consists of a guideline from i4Q^{LRT}. The **guideline** does not require digital technologies for development. Furthermore, **data analytics** and **visualisation** will use *Jupyter*. This solution is a web-based interactive development environment for code and data (**Figure 15**). This software is flexible to configure the user interface to support a wide in data science, scientific computing, and machine learning. This web uses the guideline's contents in natural language dialogs to support. Other possible solution for automatically generating document from the code is use python libraries Nbsphinx and Sphinx. Nbsphinx is a Sphinx extension that provides a source parser for *.ipynb files. Custom Sphinx directives are used to show Jupyter Notebook code in both HTML and LaTeX output. Another library is *Panda*. This library is an open-source package that allow to perform data analysis and data manipulation. It provides a fast and flexible data structures that makes it easy to work with relational and structured data.

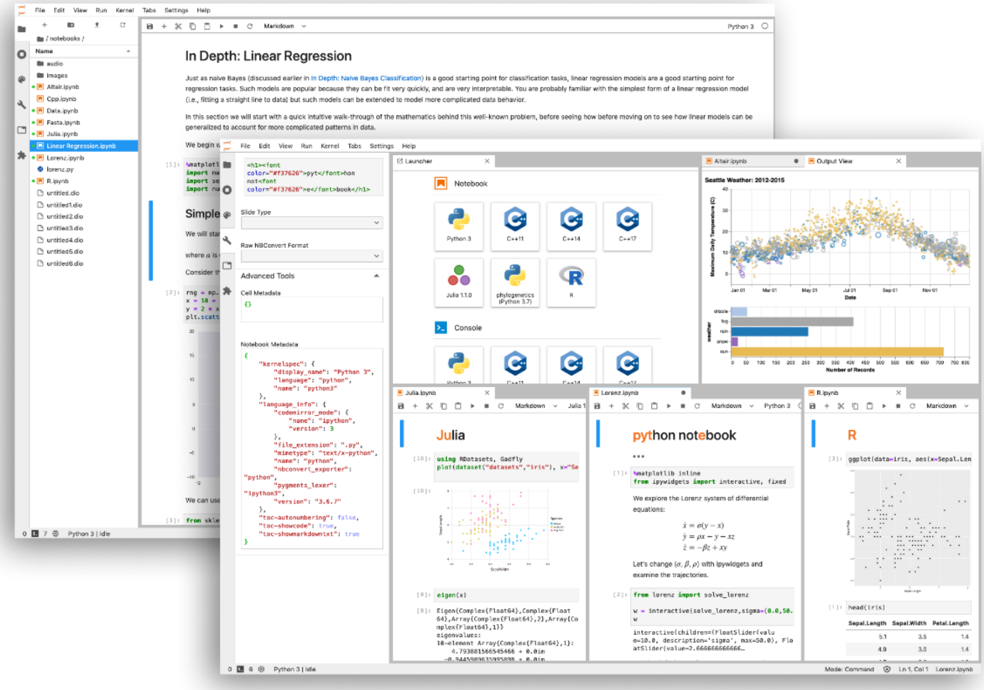

Figure 15. Jupyter Interface. (https://jupyter.org/).

5.22 i4Q Manufacturing Line Data Certification Procedure (i4Q^LCP^)

i4Q^LCP^ consists of the guideline and the digital assistant (developed by BIBA). The **guideline** does not require digital technologies for development. The SoA before and after implementation of all i4Q solutions is the basis of information to derive a certification guideline. The **digital assistant** uses the guideline’s contents in natural language dialogs to support auditors. It bases on the open-source assistant framework Rasa and builds on an infrastructure evaluated by BIBA in previous projects. Rasa uses Tensorflow.

6. Conclusions

In deliverable D1.2, a set of technologies that will be useful for the project was presented. Each technology is accompanied by the tools that can be used for its implementation and is benchmarked according to a set of features. The technologies were assigned to the solutions that will utilize them. The deliverable also contains information regarding the tools that the solution providers plan to use for the implementation of their solutions and thus can be a practical guide for the consortium partners.

More specifically, in deliverable D1.2 the state-of-the-art of emerging and promising digital technologies (e.g., Blockchain, hyperledger, fog/edge computing, data analytics, big data, machine learning, IIoT, digital twin, etc.) was characterised. Each technology was described and analysed, according to its state, maturity, tools, EU project solutions, benchmarking and assessment, application of i4Q solutions. The benchmarking evaluation framework was composed by different dimensions, divided into different criteria for the assessment of the different technologies that were analysed and selected as relevant to fulfil the development objectives of the i4Q solutions. Six dimensions were used for this analysis, namely; general, technological, business model, informational and social and various evaluation criteria, in order to characterize these technologies. Once that the different set of criteria has been defined and classified into different dimensions, each technology and associated tools were also analysed based on this evaluation framework.

Results showed that Relevance and Capability are the best-fulfilled criteria from the general criteria, while integration is the best fulfilled from the technological ones. Maturity and support are the best fulfilled from the business model criteria, need for data traceability among the informational criteria, and finally, social preferences and development-friendliness are the best-fulfilled among the social criteria. Relevance and Capability are fulfilled by all technologies analysed at the highest possible level. Furthermore, Data Analytics, Machine Learning and Big Data are the three technologies that meet most criteria, with a quite good rating. Results were used to identify the best cases – the cases in which the technologies and tools meet the most criteria, as well as which of the solutions could be used by most tools – establishing benchmarks for each of the dimensions.

Moreover, when performing the matching between the technologies and specific tools per technology and the different i4Q solutions, the results are aligned with the previous ones as Data Analytics, Machine Learning, and Big Data that are the technologies that meet most criteria with a quite good rating are also the ones in which a greater number of tools will be used for the development of the i4Q solutions. Python libraries seem to be an appropriate data analytics/visualisation related-tool for the development of 8 of the 19 i4Q solutions (in total there are 22 solutions but 5 of them are guidelines). In the case of the digital technology of machine learning, Python is the most selected tool to develop different i4Q solutions. Finally, the most important big data related-tool selected for the development of i4Q solutions is Tensorflow, as four of the six i4Q solutions that will use this technology have chosen this tool.

7. References

Abramovici, M., Göbel, J. C., & Dang, H. B. (2016). Semantic data management for the development and continuous reconfiguration of smart products and systems. *CIRP Annals*, 65(1), 185-188.

AET. (2021). Identity Lifecycle management. Retrieved from: https://www.aeteurope.com/our-company/our-business/identity-lifecycle-management/

Akselos. (2021). Akselos Solutions. Retrieved from: https://akselos.com/

Al-Jaroodi, J., & Mohamed, N. (2019). Blockchain in industries: A survey. *IEEE Access*, *7*, 36500-36515.

Alrawais, A., Alhothaily, A., Hu, C., & Cheng, X. (2017). Fog computing for the internet of things: Security and privacy issues. *IEEE Internet Computing*, 21(2), 34-42.

Andersson, (2016). Methods and Tools for Co-Simulation of Dynamic Systems with the Functional Mock-up Interface, Lund, Sweden.

Angrish, A., Craver, B., Hasan, M., & Starly, B. (2018). A case study for Blockchain in manufacturing: "FabRec": A prototype for peer-to-peer network of manufacturing nodes.*Procedia Manufacturing*,*26*, 1180-1192. 10.1016/j.promfg.2018.07.154

Bartsch, F., Neidhardt, N., Nüttgens, M., Holland, M., & Kompf, M. (2018). Anwendungsszenarien für die Blockchain-Technologie in der Industrie 4.0. *HMD Praxis der Wirtschaftsinformatik*, *55*(6), 1274-1284.

Boschert, S., & Rosen, R. (2016). Digital twin—the simulation aspect. In *Mechatronic futures* (pp. 59-74). Springer, Cham.

Boyes, H., Hallaq, B.,Cunningham, J., & Watson, T. (2018). The industrial internet of things (IIoT): An analysis framework. *Computers in industry*, 101, 1-12.

Bsquare (2017). 2017 Annual IioT Maturity Survey: Adoption of IioT in Manufacturing, Oil and Gas and Transportation. Retrieved from: https://3q0ds8402hawyzjwb3qrnh43-wpengine.netdna-ssl.com/wp-content/uploads/2018/01/2017bsquareIIoT.pdf

BusinessWire, (2020). REPLY: AXULUS Makes IioT Projects Scalable. Retrieved from: https://www.businesswire.com/news/home/20201119005209/en/REPLY-AXULUS-Makes-IIoT-Projects-Scalable

C2NET: D3.3. (2016) 1st Data collection from resources virtualization. Retrieved from: http://c2net-project.eu/deliverables/-/blogs/d3-3-1st-data-collection-from-resources-virtualization

Cachin, C. (2016, July). Architecture of the Hyperledger blockchain fabric. In *Workshop on distributed cryptocurrencies and consensus ledgers* (Vol. 310, No. 4).

CheManager. (2020). An Open Framework for Infrastructure Digital Twins Open Source and Collaborative Models Build Trust, Reduce Costs, and Create Value. Retrieved from: https://www.chemanager-online.com/en/topics/information-technology/open-framework-infrastructure-digital-twins

Confais, B., Lebre, A., & Parrein, B. (2016, December). Performance analysis of object store systems in a fog/edge computing infrastructures. In *2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)* (pp. 294-301). IEEE.

CORDIS. (2021). EU – Research Results. Retrieved from European Commission: https://cordis.europa.eu/

CREMA: D4.1(2015). CREMA Data Model, Model Library and Profiles Prototype I. Retrieved from: https://www.crema-project.eu/media/1087/t41-d41-data-model-library-profiles-prototpye-i.pdf

Dassault Systèmes. (2018).How Digital Twin Tech Helps Drive Industry. Retrieved from: https://www.forbes.com/sites/dassaultsystemes/2018/11/13/how-digital-twin-tech-is-helping-drive-industry/?sh=1902cad43b0c

Eclipse. (2021). Ditto. where IoT devices and their digital twins get together. Retrieved from: https://www.eclipse.org/ditto/index.html

Elias, A. (2021). Understanding Hardware Roots of Trust. Retrieved from: https://www.synopsys.com/designware-ip/technical-bulletin/understanding-hardware-roots-of-trust-2017q4.html

Emmrich, V., Döbele, M., Bauernhansl, T., Paulus-Rohmer, D., Schatz, A., & Weskamp, M. (2015). Geschäftsmodell-Innovation durch Industrie 4.0: Chancen und Risiken für den Maschinen- und Anlagenbau. München: Wieselhuber & Partner, Fraunhofer IPA.

Erpinnews (2018). Fog computing vs edge computing. Retrieved from: https://erpinnews.com/fog-computing-vs-edge-computing

Faisal, M., Ali, I., Khan, M. S., Kim, S. M., & Kim, J. (2020). Establishment of Trust in Internet of Things by Integrating Trusted Platform Module: To Counter Cybersecurity Challenges. *Complexity*, *2020*.

Gartner (2019). Gartner's June 2019 Magic Quadrant for Industrial IoT Platform. Retrieved from Gartner Magic Quadrant: https://www.gartner.com/en/research/methodologies/magic-quadrants-research

Gatto, J. (2021). What are the Risks Associated with Industrial IoT (Industrial Internet of Things)? Retrieved from Attila Security: https://www.attilasec.com/blog/what-are-the-risks-associated-with-industrial-iot

Gedeon, J., Brandherm, F., Egert, R., Grube, T., & Mühlhäuser, M. (2019). What the fog? edge computing revisited: Promises, applications and future challenges. *IEEE Access*, *7*, 152847-152878.

GE-Digital. (2021) Digital Twins are mission critical. Retrieved from General Electric: https://www.ge.com/digital/applications/digital-twin

Groggert, S., Wenking, M., Schmitt, R. H., & Friedli, T. (2017, December). Status quo and future potential of manufacturing data analytics—an empirical study. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 779-783). IEEE.

Hassan, N., Gillani, S., Ahmed, E., Yaqoob, I., & Imran, M. (2018). The role of edge computing in internet of things. *IEEE Communications Magazine*, 56(11), 110-115.

Heck, M., Edinger, J., Schäfer, D., & Becker, C. (2018, July). IoT applications in fog and edge computing: Where are we and where are we going? *In 2018 27th International Conference on Computer Communication and Networks (ICCCN)* (pp. 1-6). IEEE.

Helmiö, P. (2018). Open source in industrial Internet of Things: A systematic literature review. Master's Thesis, School of Business and Management. Lappeenranta University of Technology, p. 21

Hong, C. H., & Varghese, B. (2019). Resource management in fog/edge computing: a survey on architectures, infrastructure, and algorithms. *ACM Computing Surveys (CSUR)*, 52(5), 1-37.

Hu, P., Dhelim, S., Ning, H., & Qiu, T. (2017). Survey on fog computing: architecture, key technologies, applications and open issues. *Journal of network and computer applications*, *98*, 27-42.

IET. (2021). Digital twins for the built environment. An introduction to the opportunities, benefits, challenges and risks. Retrieved from Institution of Engineering and Technology: theiet.org/digital-twins

IIoT-World. (2021). Four most hard to solve IioT security issues. Retrieved from: https://iiot-world.com/ics-security/cybersecurity/four-most-hard-to-solve-iiot-security-issues/

INCIBE (2015). IEC 62443: Evolución de la ISA 99. Retrieved from: https://www.incibe-cert.es/blog/iec62443-evolucion-isa99

ISA. (2021). New ISA/IEC 62443 standard specifies security capabilities for control system components. Retrieved from: https://www.isa.org/intech-home/2018/september-october/departments/new-standard-specifies-security-capabilities-for-c

iTwinsjs. (2021). A starter kit for developing web applications for infrastructure digital twins. Retrieved from: https://www.itwinjs.org/

Keenan, M. (2020). Towards Proof-Of-Value: A Tale of IIoT Maturity. Retrieved from: https://www.iotforall.com/iiot-maturity

Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. (2019). Edge computing: A survey. *Future Generation Computer Systems*, 97, 219-235.

Kosan, L. (2019). Top 10 IioT Platforms. Retrieved from IoT World Today: https://www.iotworldtoday.com/2019/08/07/top-10-iiot-platforms/

Kraft, E. M. (2016). The air force digital thread/digital twin-life cycle integration and use of computational and experimental knowledge. In 54th *AIAA aerospace sciences meeting* (p. 0897).

LeCun, Y.,Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Lengthorn, P. (2021). How much to build a Digital Twin? Retrieved from: https://www.consultengsurvivor.com/cost-to-build-a-digital-twin

Liu, L. and Chi, L., 2002. Evolutional Data Quality: A Theory-Specific View. In *ICIQ* (pp. 292-304).

Lohmer, J. (2019). Applicability of blockchain technology in scheduling resources within distributed manufacturing. In Logistics management. *Proceedings of the German academic association for business research 2019* (pp. 89–103). Springer.

Lohmer, J., & Lasch, R. (2020). Blockchain in operations management and manufacturing: Potential and barriers. *Computers & Industrial Engineering*, 149, 106789.

M. Gonzalez, Virtual Sensor Approaches for Condition Monitoring of Vertical Transportation Systems, PhD Thesis, KU Leuven, 2020.

Microsoft. (2021). Azure Digital Twins. Retrieved from: https://azure.microsoft.com/en-us/services/digital-twins/

Musketeer: D4.2. (2019). Pre-processing, normalization, data alignment and data value estimation algorithms –v1. Retrieved from: https://musketeer.eu/wp-content/uploads/2019/10/MUSKETEER_D4.2.pdf

Negri, E., Fumagalli, L., & Macchi, M. (2017). A review of the roles of digital twin in CPS-based production systems. *Procedia Manufacturing*, 11, 939-948.

Nofer, M., Gomber, P., Hinz, O., & Schiereck, D. (2017). Blockchain. *Business & Information Systems Engineering,* 59(3), 183-187.

Parrot, A., Warshaw, L. (2017). Industry 4.0 and the digital twin. Deloitte Insights. Retrieved (2021-02-17) from: https://www2.deloitte.com/us/en/insights/focus/industry-4-0/digital-twin-technology-smart-factory.html

Perino, J. (2019). 6 Most Asked Questions from the Digital Twin for Process Industry Webcast. Retrieved from Industrial Transformation Blog: https://blog.lnsresearch.com/6-most-asked-questions-from-digital-twin-process-industry-webcast

Platenius-Mohr, M., Malakuti, S., Grüner, S., & Goldschmidt, T. (2019, October). Interoperable Digital Twins in IIoT Systems by Transformation of Information Models: A Case Study with Asset Administration Shell. In *Proceedings of the 9th International Conference on the Internet of Things* (pp. 1-8).

Porter, M. E., & Heppelmann, J. E. (2015). How smart, connected products are transforming companies. *Harvard business review*, 93(10), 96-114.

RAMBUS (2021). Hardware Root of Trust. Retrieved from: https://www.rambus.com/blogs/hardware-root-of-trust/

Robins, B. (2020). Digital Twins: The Path to Maturity. Retrieved from Bentley Systems: https://www.bentley.com/en/perspectives-and-viewpoints/topics/perspectives/2020/digital-twins-path-to-maturity

Rosen, R., Von Wichert, G., Lo, G., & Bettenhausen, K. D. (2015). About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine*, 48(3), 567-572.

RTI (2015). Industrial Internet of Things Real Time Innovations. Retrieved from: https://info.rti.com/hubfs/docs/Industrial_IoT_FAQ.pdf.

Sandner, P. (2017). Application of blockchain technology in the manufacturing industry. *Frankfurt School Blockchain Center.*

Schluse, M., & Rossmann, J. (2016, October). From simulation to experimentable digital twins: Simulation-based development and operation of complex technical systems. In *2016 IEEE International Symposium on Systems Engineering (ISSE*) (pp. 1-6). IEEE.

Schutt, R. O'Neil, C. (2013). Doing Data Science.O'Reilly Media. ISBN 978-1-449-35865-5.

Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., & Wang, L. (2012). Modeling, simulation, information technology & processing roadmap. *National Aeronautics and Space Administration*.

Shpantzer, G. (2013). Implementing Hardware Roots of Trust: The Trusted Platform Module Comes of Age. Retrieved from: https://trustedcomputinggroup.org/wp-content/uploads/SANS-Implementing-Hardware-Roots-of-Trust.pdf

Siemens (2021). Digitalization in industry: Twins with potential. Retrieved from Siemens: https://new.siemens.com/global/en/company/stories/industry/the-digital-twin.html

Socha, D. (2018). The digital twin maturity continuum. Retrieved from Smart Energy International: https://www.smart-energy.com/industry-sectors/policy-regulation/digital-twins-maturity-continuum-david-socha/

Steer, M. (2018). Will there be a Digital Twin for Everything and everyone? Retrieved from: https://www.digitalistmag.com/iot/2018/05/23/will-there-be-digital-twin-for-everything-everyone-06169041/

Susila, N., Sruthi, A., & Usha, S. (2020). Impact of cloud security in digital twin. In *Advances in Computers* (Vol. 117, No. 1, pp. 247-263). Elsevier.

Tadapaneni, N. R. (2019). Role of fog computing in the internet of things. *International Journal of Scientific Research and Engineering Trends*.

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 94(9), 3563-3576

Tuegel, E. J., Ingraffea, A. R., Eason, T. G., & Spottswood, S. M. (2011). Reengineering aircraft structural life prediction using a digital twin. *International Journal of Aerospace Engineering*.

Tune, N. (2021) Beyond buzzwords: digital twin maturity spectrum. Retrieved from SNC-Lavalin: https://www.snclavalin.com/en/beyond-engineering/beyond-buzzwords-digital-twin-maturity-spectrum

Van Kruijsdijk, K. (2018). Digital Twins as R&D Accelerators – The Case for an Open Source Ecosystem. Shell Global Solutions International. Retrieved from: https://ercim-news.ercim.eu/en115/special/2096-digital-twins-as-r-d-accelerators-the-case-for-an-open-source-ecosystem

Wang, X., Han, Y., Leung, V. C.,Niyato, D., Yan, X., & Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(2), 869-904.

Wikipedia (2021). Digital Twins. Retrieved from: https://en.wikipedia.org/wiki/Digital_twin

Winther, C. (2020). Modelon. Retrieved from: https://www.modelon.com/co-simulation-using-the-open-source-python-package-pyfmi/.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data.*IEEE transactions on knowledge and data engineering*, *26*(1), 97-107.

Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2019). Blockchain technology overview. *arXiv preprint arXiv:1906.11078*.

Annex I

	i4Q Solutions																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	$i4Q^{QE}$	$i4Q^{DQG}$	$i4Q^{BC}$	$i4Q^{TN}$	$i4Q^{SH}$	$i4Q^{CSG}$	$i4Q^{DR}$	$i4Q^{DRG}$	$i4Q^{DIT}$	$i4Q^{DA}$	$i4Q^{BDA}$	$i4Q^{AD}$	$i4Q^{AI}$	$i4Q^{EW}$	$i4Q^{IM}$	$i4Q^{DT}$	$i4Q^{PQ}$	$i4Q^{QD}$	$i4Q^{PA}$	$i4Q^{LRT}$	$i4Q^{LRG}$	$i4Q^{LCP}$
Blockchain/Hyperledger			X																			
Hyperledger Fabric			X																			
Solidity																						
Geth																						
Solc																						
Mist																						
Remix																						
Truffle																						
Metamask																						
Ganache																						
Blockchain testnet																						
Fog/Edge Computing																						
AWS IoT Greengrass																						
AWS IoT SiteWise																						
IBM Edge Application Manager													X	X								
ZDMP Distributed computing																						
Bosch IoT Edge																						
Microsoft Azure IoT Edge																						
Oracle Roving Edge Infrastructure																						
FogAtlas																						
OpenVolcano																						
ThingWorx Kepware Edge																						
KubeEdge													o	o								
EdgeXFounDry													o	o								
AKRAINO																						
StarlingX																						
OPNFV													o	o								
Eclipse Foundation																						
ETSI																						
Open Connectivity Foundation																						
ITU																						
Nerve Blue													X	X								
OpenFog																						
Open Horizon													X	X								

	i4Q Solutions																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	$i4Q^{QE}$	$i4Q^{DQG}$	$i4Q^{BC}$	$i4Q^{TN}$	$i4Q^{SH}$	$i4Q^{CSG}$	$i4Q^{DR}$	$i4Q^{DRG}$	$i4Q^{DIT}$	$i4Q^{DA}$	$i4Q^{BDA}$	$i4Q^{AD}$	$i4Q^{AI}$	$i4Q^{EW}$	$i4Q^{IM}$	$i4Q^{DT}$	$i4Q^{PQ}$	$i4Q^{QD}$	$i4Q^{PA}$	$i4Q^{LRT}$	$i4Q^{LRG}$	$i4Q^{LCP}$
Data Analytics/visualisation									X			X			X	X	X		X	X	X	
Tableau Software																						
Microsoft azure																						
Python libraries									X			X			X	X	X		X	X	X	
R libraries									X						X	o	X		o			
Grafana									x			X				o			o			
Jupyter									X			X				o			o		X	
Apache Superset												X										
Big Data	X									X	X				X		X	X				
Microsoft Azure																						
Tensorflow	X									X	X						X					
Keras										X							X					
FI-WARE Orion Context Broker																						
Apache Flink										X	X											
Apache Storm										X	X											
Apache Spark										X	X											
Machine Learning									X	X					X	X	X	X	X			
Apache Spark MLlib									X	X					X							
R									X							o	X		o			
WEKA										X												
Python									X	X					X	X	X		X	X		
Mahout																						

	i4Q Solutions																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	$i4Q^{QE}$	$i4Q^{DQG}$	$i4Q^{BC}$	$i4Q^{TN}$	$i4Q^{SH}$	$i4Q^{CSG}$	$i4Q^{DR}$	$i4Q^{DRG}$	$i4Q^{DIT}$	$i4Q^{DA}$	$i4Q^{BDA}$	$i4Q^{AD}$	$i4Q^{AI}$	$i4Q^{EW}$	$i4Q^{IM}$	$i4Q^{DT}$	$i4Q^{PQ}$	$i4Q^{QD}$	$i4Q^{PA}$	$i4Q^{LRT}$	$i4Q^{LRG}$	$i4Q^{LCP}$
IIoT				X																		
Hardware				X																		
Arduino																						
BeagleBoard				X																		
Microduino																						
Node MCU (ESP 8266)																						
OLinuXino																						
Raspberry				X																		
Tessel																						
OpenMote B				X																		
Zolertia RE-Mote				X																		
Software, operating systems			X	X			X						X	X								
Contiki				X																		
OpenWrt																						
Tiny OS																						
Linux			X	X			X						X	X								
SDNWise				X																		
Open Network Operatin System (ONOS)				X																		
Ditto project																X						
iModel.js																						
PyFMI																X						
FMI compliant simulation Tool (i.e. OpenModelica)																X						
Security Technologies					X	X																
62443 standard					X	X																
Digital identity life cycle (asymmetric cryptography)					X	X																
Hardware root of trust					X	X																