

Tailored IoT & BigData Sandboxes and Testbeds for Smart, Autonomous and Personalized Services in the European Finance and Insurance Services Ecosystem

# Infinitech

## D2.5 – Specifications of INFINITECH Technologies- I

<b>Lead Beneficiary</b>	<b>GFT</b>
<b>Due Date</b>	2020-05-31
<b>Delivered Date</b>	2020-05-31
<b>Revision Number</b>	2.0
<b>Dissemination Level</b>	Public (PU)
<b>Type</b>	Report (R)
<b>Document Status</b>	Draft
<b>Review Status</b>	Quality Assurance Reviewed
<b>Document Acceptance</b>	WP Leader Accepted and Coordinator Accepted
<b>EC Project Officer</b>	Pierre-Paul Sondag

HORIZON 2020 - ICT-11-2018



This project has received funding from the European Union's horizon 2020 research and innovation programme under grant agreement no 856632

## Contributing Partners

Partner Acronym	Role <sup>1</sup>	Name Surname <sup>2</sup>
<b>GFT</b>	Contributor	Maurizio Megliola
<b>GRAD</b>	Contributor	Lilian Adkinson Orellana
<b>LXS</b>	Contributor	Patricio Martinez
<b>LXS</b>	Contributor	Jacob Roldan
<b>LXS</b>	Contributor	Sandra Ebro
<b>LXS</b>	Contributor	Jose María Zaragoza
<b>Atos</b>	Contributor	Ignacio Elicegui Maestro Juan Carrasco José Gato Nuria Ituarte Aranda
<b>INNOV</b>	Contributor	John Soldatos
<b>ORT GFT ENG UNP NUIG JSI AKITF BOUN UPRC RB</b>	Contributor	
<b>UNP</b>	Reviewer	
<b>CTAG</b>	Reviewer	
<b>INNOV</b>	QA	

## Revision History

Version	Date	Partner(s)	Description
<b>0.1</b>	2019-12-31	ORT	ToC Version
<b>0.2</b>	2020-05-22	GFT, ALL	First Version for Internal Review
<b>1.0</b>	2020-05-28	GFT, ALL	Version for Quality Assurance
<b>2.0</b>	2020-05-31	GFT	Version for Submission

<sup>1</sup> Lead Beneficiary, Contributor, Internal Reviewer, Quality Assurance

<sup>2</sup> Can be left void

## Executive Summary

The Deliverable D2.5 of the INFINITECH project is a fundamental part of the process that leads to the specification of the technologies that constitute the **building blocks** that will be utilized by the pilots.

As part of the Horizon 2020 Programme Innovation Action, the INFINITECH project has the ambition to advance the state of the art of the technology integrating and developing the components needed to demonstrate the feasibility of the different demonstrators. Moreover, the developed solutions will allow to approach a vast class use cases related to data processing within the **Financial and Insurance Sectors**.

To advance the actual status of development an assessment of what is available and what is desirable is needed and this Deliverable has this objective.

In particular, the deliverable lists the tools and technologies currently available and in development by the technology partners. The deliverable also contains detailed specifications of the components identified by the pilots, including Input and Output formats, functionalities and specifications about the implementation technologies (e.g., BigData/IoT platforms, AI/ML toolkits, HPC infrastructures) that will be used to realize them. The specifications take into account the requirements and specifications of the previous two tasks of the project, namely T2.1 and T2.2 that deals with User Stories and Reference Scenarios, while at the same time aligning to the INFINITECH-RA that will be developed as part of T2.7.

The deliverable is also part of a larger picture as it also provides the basis for future work and material for the **Marketplace** and the **Virtualized Digital Innovation Hub** of WP8.

Rather than reinventing the wheel, the INFINITECH project will build on the shoulders of the giants of the many and different assets brought from the technological partners and fintech companies as well as the assets provided by end-users and stakeholders involved in the project. In the course of action, the project will explore other ways to integrate state-of-the-art tools and technologies that will fulfil the components needed to solve a larger class of problems in the Financial and Insurance domain. Future versions of this deliverable will address this broader scope.

## Table of Contents

1.	Introduction .....	9
1.1.	Objective of the Deliverable .....	9
1.2.	Insights from other Tasks and Deliverables .....	9
1.3.	Structure .....	9
2.	Available Tools and Technologies .....	10
	Background technology .....	18
3.	INFINITECH components specifications .....	24
	Template for Data Component Description.....	25
3.1	Component Group: Data Management .....	26
3.1.1	Relational Database .....	26
3.1.2	Polyglot Query Processing .....	26
3.1.3	Incremental Analytics.....	27
3.1.4	Filtering .....	28
3.1.5	Joiner .....	29
3.1.6	OneHotEncoder.....	30
3.1.7	Visualization Preparation.....	30
3.1.8	Stream Processor .....	31
3.2	Component Group: Security and Privacy.....	32
3.2.1	eIDAS integration .....	32
3.2.2	Data Anonymization.....	33
3.2.3	Digital User Onboarding System (DUOS) .....	34
3.2.4	Data Protection Orchestrator (DPO).....	36
3.3	Component Group: Blockchain and Information Sharing.....	36
3.3.1	Blockchain Reader.....	36
3.3.2	Blockchain Writer.....	37
3.3.3	Smart Contract Executor .....	38
3.3.4	Blockchain Data Visualizer .....	40
3.3.5	Blockchain Authenticator.....	41
3.3.6	Blockchain Encryptor .....	42
3.3.7	Blockchain Decryptor .....	43
3.3.8	Blockchain Transaction Dataset Preparation Component.....	43
3.3.9	Scalable Transaction Graph Analysis Component.....	44
3.3.10	User Interface for Blockchain Transaction Reports and Visualization Component .....	45
3.4	Component Group: Data Semantics .....	47
3.4.1	Semantic Stream Analyzer .....	47
3.4.2	Semantic Reasoner .....	47

## D2.5 – Specifications of INFINITECH Technologies - I

3.4.3	Ontology Mapping .....	47
3.4.4	Preprocessing.....	47
3.4.5	Anomaly Analysis .....	47
3.4.6	Pattern Analysis .....	48
3.4.7	Stream Story.....	48
3.5	Component Group: Internet of Things .....	49
3.5.1	IoT Registry.....	49
3.5.2	IoT Stream Handler .....	50
3.6	Component Group: Machine Learning .....	51
3.6.1	Fraud Detection Service Training.....	51
3.6.2	Fraud Detection Service Execution .....	52
3.6.3	Pay As You Drive Service Training .....	53
3.6.4	Pay As You Drive Service Execution .....	53
3.6.5	Recommendation Engine Training.....	54
3.6.6	Recommendation Engine Execution .....	54
3.6.7	Association Rules Training .....	54
3.6.8	Association Rules Execution.....	54
3.6.9	Unexpected transaction detector .....	54
3.6.10	Account transaction forecaster.....	55
3.6.11	Overdraft forecaster .....	56
3.6.12	Automation proposer.....	57
3.6.13	Transaction sentinel.....	57
3.6.13	Bank’s product recommender .....	58
3.6.14	Smart expense advisor .....	59
3.6.15	Savings proposer .....	60
3.6.16	Recommender.....	60
3.6.17	Cash Flow Prediction Model .....	61
3.6.18	Budget Prediction .....	63
3.6.19	KPI Engine.....	64
3.6.20	Transaction (Txn) Monitoring .....	66
3.6.21	Transaction (Txn) Categorization .....	67
3.6.22	Invoice Processing.....	68
3.6.23	Benchmark(s) .....	70
3.6.24	KMeans .....	71
3.6.25	Random Forest (Model) .....	72
3.6.26	Random Forest (Predict) .....	73
3.7	Component Group: Visualization.....	74
3.7.1	Data Visualization Dashboard .....	74

## D2.5 – Specifications of INFINITECH Technologies - I

3.7.2	Model Visualization Dashboard .....	74
3.7.3	Results Visualization Dashboard .....	74
3.7.4	Real-Time Visualization .....	74
4.	Conclusions and next steps.....	75

### List of Figures

Figure 1 – Tasks interactions in WP2 .....	9
Figure 2 – Logical Schema of Data Processing Components in INFINITECH RA.....	24

### List of Tables

Table 1 – Tools and Technology Assessment.....	10
Table 2 – Background technology.....	18
Table 3 - Component Description Template .....	25

## Abbreviations

<b>Acronym</b>	<b>Definition</b>
<b>3DS</b>	Three-Domain Secure
<b>AAA</b>	Authentication Authorization Accounting
<b>ACID</b>	Atomicity, Consistency, Isolation, Durability
<b>ADC</b>	Account Data Compromise
<b>AML</b>	Anti Money Laundering
<b>API</b>	Application Programming Interface
<b>BAIT</b>	Bankaufsichtliche Anforderungen an die IT
<b>BCM</b>	Business Continuity Management
<b>CEBS</b>	Committee of European Banking Supervisors
<b>CIA</b>	Confidentiality Integrity and Availability
<b>CJEU</b>	Court of Justice of the European Union
<b>CMDB</b>	Configuration Management Database
<b>CSC</b>	Common and Secure Communication
<b>CSV</b>	Comma Separated Value files
<b>DaaS</b>	Data-as-a-Service
<b>DLT</b>	Distributed ledger technology
<b>DoW</b>	Description of Work
<b>DPA</b>	Data Protection Authority
<b>DPF</b>	Data Policy Framework
<b>DPO</b>	Data Protection Officer
<b>DSP</b>	Digital Service Providers
<b>EBA</b>	European Banking Authority
<b>ECB</b>	European Central Bank
<b>eID</b>	electronic Identification
<b>eIDAS</b>	electronic IDentification Authentication and trust Services
<b>eTS</b>	electronic Trust Services
<b>EU</b>	European Union
<b>EUC</b>	End-User Computing
<b>FATCA</b>	Foreign Account Tax Compliance Act
<b>FLOSS</b>	Free/Libre Open Source Software
<b>GDPR</b>	General Data Protection Regulation
<b>GLB</b>	Gramm-Leach-Bliley Act
<b>H2020</b>	Horizon 2020 Programme
<b>HTML</b>	Hypertext Markup Language
<b>IaaS</b>	Infrastructure as a Service
<b>ICT</b>	Information Communication Technologies
<b>IEC</b>	International Electrotechnical Commission
<b>IRA</b>	INFINITECH Reference Architecture
<b>ISMS</b>	Information Security Management System
<b>ISO</b>	International Organization for Standardization
<b>ISP</b>	Internet Service Provider

<b>JSON</b>	JavaScript Object Notation
<b>JWT</b>	JSON Web Token
<b>KYC</b>	Know Your Customer
<b>LOPD</b>	Law for Protection of Personal Data
<b>MiFID</b>	Markets in Financial Instruments Directive
<b>MiFIR</b>	Markets in Financial Instruments and Amending Regulation
<b>NDA</b>	Non-Disclosure Agreement
<b>NIS</b>	Network and Information Systems
<b>NLP</b>	Natural language processing
<b>OES</b>	Operators of Essential Services
<b>OSS</b>	Open Source Software
<b>P2PP</b>	Peer-to-Peer Payment
<b>PaaS</b>	Platform as a Service
<b>PAN</b>	Primary Account Number
<b>PCI-DSS</b>	Payment Card Industry Data Security Standard
<b>PIA</b>	Privacy Impact Assessment
<b>PM</b>	Person Month
<b>PSD2</b>	Payment Service Directive 2
<b>PSP</b>	Payment Service Provider
<b>PSPS</b>	Public Safety and Personal Security
<b>PSU</b>	Payment Service User
<b>PU</b>	Public
<b>QTSP</b>	Qualified Trust Service Provider
<b>RA</b>	Reference Architecture – IRA INFINITECH Reference Architecture
<b>RDF</b>	Resource Description Framework
<b>REST</b>	Representational State Transfer
<b>RTD</b>	Research and Development
<b>RTS</b>	Regulatory Technical Standard
<b>SA</b>	Supervisory Authority
<b>SCA</b>	Strong Customer Authentication
<b>SECaaS</b>	Security-as-a- Service
<b>SME</b>	Small and Medium-Sized Enterprises
<b>SQL</b>	Structured Query Language
<b>SSL</b>	Secure Sockets Layer
<b>TLS</b>	Transport Layer Security
<b>UI</b>	User Interface
<b>URL</b>	Uniform Resource Locator
<b>XML</b>	Extensible Markup Language



## 1. Introduction

The scope of this section is to introduce the deliverable and familiarize the user with its contents. Towards this end, the current section summarises the objective of the current deliverable, its relation to the other work packages and corresponding deliverables and analyses its structure.

### 1.1. Objective of the Deliverable

The objectives and scope of this deliverable are to detail the BigData/IoT technological building blocks that will be developed in the scope of the project, in the areas of data and semantic interoperability, cost-effective real-time BigData analytics, elastic cloud storage, integrated (declarative) data querying, AI/ML algorithms and more.

### 1.2. Insights from other Tasks and Deliverables

The specifications included in this deliverable take into account the requirements and specifications of the previous two tasks [1] and [2], while at the same time aligning to the INFINITECH-RA that will be developed as part of T2.7.

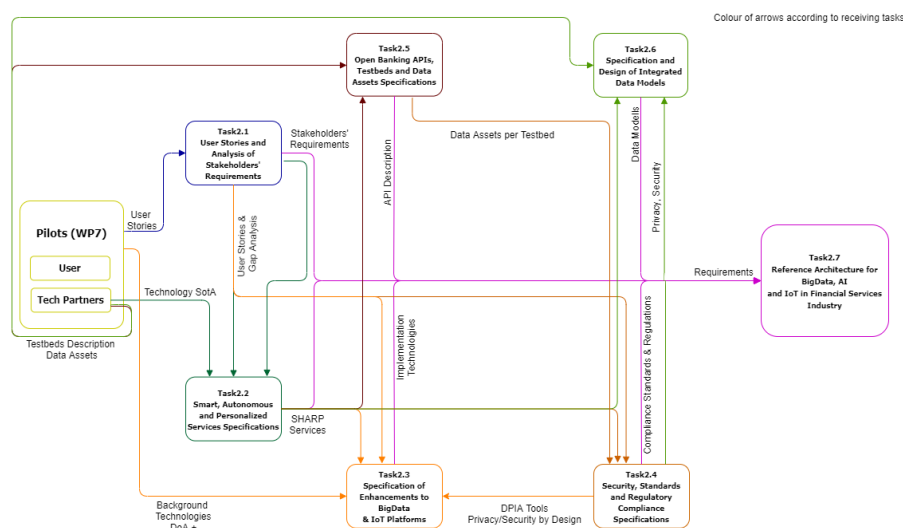


Figure 1 – Tasks interactions in WP2

### 1.3. Structure

This deliverable is composed of four main sections. Chapter 1 is the introduction to the deliverable and includes the description of the objective, insights from other tasks and deliverables and the structure. Chapter 2 describes in detail the available technologies within the context of the project. Chapter 3 details the specifications of the INFINITECH components. Finally, Chapter 4 reports some conclusions.

## 2. Available Tools and Technologies

The following table describes available tools and technologies to be used in the INFINITECH platform, identified after a process/reasoning based on the input provided by the Consortium Partners.

For each tool or technology, the following basic information are provided:

- Company ACRONYM owner of the tool/platform;
- Name/Short Description of tool/platform;
- Technology Readiness Level current and expected level at the end of the INFINITECH project;
- Characteristics/Technology;
- Reference to Documentation/links/Demo environment;
- Ideas for enhancements to be done in the course of the project;
- Targeted pilots (pilots tools/platform aim at); and finally, the Availability of interfaces/API.

Table 1 – Tools and Technology Assessment

Company	Name of Tool / Platform	TRL	Characteristics / Technology	Reference to Documentation / Links / Demo environment	Ideas for enhancements	Targeted pilots (pilots tools / platform aim at)	Availability of interfaces / API
<b>ATOS</b>							
a)	Service Provider eIDAS Integration (SPEiDI)	6->7	This technology enables connecting online services with eIDAS infrastructure to European eID use. It allows authentication of customer against the pan European eID infrastructure. It will be one of the data governance building blocks of the platform (in WP3). It will be used in pilots involving customer authentication and cross-border transactions (e.g., ARI, BOI and SIA led pilots).	LEPS project: D5.2. Develop an eIDAS Interconnection Supporting Service <a href="http://www.leps-project.eu/node/351">http://www.leps-project.eu/node/351</a>	This connector requires the use of eID schemas under eIDAS <a href="https://ec.europa.eu/cefdigital/wiki/display/EIDCOMMUNITY/Overview+of+pre-notified+and+notified+eID+schemes+under+eIDAS">https://ec.europa.eu/cefdigital/wiki/display/EIDCOMMUNITY/Overview+of+pre-notified+and+notified+eID+schemes+under+eIDAS</a> It is required the use of eIDs valid for countries that notified their participation in eIDAS network There will be implemented the needed improvements regarding		

D2.5 – Specifications of INFINITECH Technologies - I

					the communication (protocol) of SPeIDI with SP and eIDAS node		
<b>b)</b>	Car data ingestion	5->7	This technology is tasked with the capture, homogenization, distribution and storage of the datasets that support the connected car pilot. It involves the design and implementation of the IoT agents that adapt the data available from the car, as well as deploying the necessary modules for data storage and distribution (FIWARE Orion, FIWARE Cygnus, etc)		As the data sources size increases, the technology will have to become more robust by polishing the performance of the deployed IoT agents.		
<b>c)</b>	Data Protection Orchestrator (DPO)	5-> 6	It is an enabler for embedding and automating the assurance of security and privacy by design and by default in heterogeneous and complex business flow. It orchestrates various privacy and security management functions (such as access control, encryption and anonymization). It will be used as part of the data governance Framework of the Project, and towards establishing the regulatory compliance tools in the project's sandboxes.	Witdom Project: <a href="http://witdom.eu/sites/default/files/witdom/public/content-files/deliverables/D4.4_Final%20specification%20of%20system%20architectures%20for%20WITDOM%20scenarios_v1.0_final_20171031_0.pdf">http://witdom.eu/sites/default/files/witdom/public/content-files/deliverables/D4.4_Final%20specification%20of%20system%20architectures%20for%20WITDOM%20scenarios_v1.0_final_20171031_0.pdf</a>	It requires Swagger specification of the components (PETs) that will be called via REST by DPO There will be developed the business flows to address the specific communication with the components		
<b>d)</b>	Digital User Onboarding System (DUOS)	5-> 6	It is a solution for dealing with virtual identities in a mobile device. It provides remote user registration using eID or passport	ARIES project: <a href="https://www.aries-project.eu/content/aries-prototype-instantiation">https://www.aries-project.eu/content/aries-prototype-instantiation</a> ,	It is needed to use eIDs issued by European National authorities according to the EU eID schemas: eID cards and Passports There will be implemented the needed improvements regarding integration with end users application (Bank application) that needs user authentication		
<b>e)</b>	EASIER-AI	4->6	EASIER-AI is an Hybrid (Cloud/Edge) platform that facilitates to develop, measure, monitor and deploy your AI models. The platform facilitates the data science tasks and it is focused on working on Hybrid Infrastructure and exploiting data generated by IoT. The platform synchronizes Cloud and Edge, keeping the Edge		By including this tool in the INFINITECH project, we aim to feed it with new dataset sources, resulting in the development of new ML models for the platform.		

D2.5 – Specifications of INFINITECH Technologies - I

			always up to date to run always the most accurate model.				
<b>f)</b>	Driver profile classifier	5->7	This technology is aimed at the use of high-quality vehicle data allows insurance companies to offer customized products. The application of supervised machine learning techniques is proposed to classify drivers' profiles which generates a customized insurance premium. The resulting model is then deployed with TensorFlow serving and integrated as part of the cloud platform with a wrapper, achieving an accuracy of 85.7%.		The inclusion of this model in the project could bring an improvement on the data analysis of the datasets, bringing an improvement to the model accuracy.		
<b>e)</b>	Distributed near-real-time HPC processing and exchange of IoT streaming data	5->6	IoT & AI & HPC				
<b>CP</b>							
<b>a)</b>	Botakis Chatbot Development Network	7->8	A tool for rapid development of chatbots applications, which will be used for the development of chatbots features in the INFINITECH pilots (i.e. notably the LIB, BOC and NBG led pilots).	<a href="https://www.crowdpolicy.com/service/botakis/">https://www.crowdpolicy.com/service/botakis/</a>	Enhancements expecting to be achieved for Botakis Chatbot Platform, based on INFINITECH pilots (i.e. notably the GFT and NBG led pilots) : - Built-in dialogs that utilize and be integrated with existing NLP frameworks (open or proprietary) provided by partners or every interested party - Powerful dialog system with dialogs that are isolated and composable. - Built-in prompts for simple things like Yes/No, strings, numbers, enumerations.	Pilot 5a, Pilot 5b	Botakis Platform provides an API in order to be integrated with other systems and provide any relative data through the chatbot functionality that is available. Also is already integrated with specific public available web resources, to

D2.5 – Specifications of INFINITECH Technologies - I

							provide generic info regarding Weather Data, Astrology, Local News, etc.
<b>b)</b>	Crowdpolicy Open (innovation) banking solution	6/7->8/9	Crowdpolicy Open (innovation) banking platform is as set of predefined and customisable banking web services and data models integrated with our own API Manager that supports access control, monitoring and authentication. Our solution puts the bank (or any monetary financial institution) in control of the third-party partner relation. The solution is full PSD2 & GDPR Compliance	<a href="https://www.crowdpolicy.com/fintech-innovation-enabler/">https://www.crowdpolicy.com/fintech-innovation-enabler/</a>	Enhancement aim through INFINITECH project are : - technology scale-up is to democratise the use and exploitation of open banking APIs even for users with no development skills, building fintech software development kits. - implement a complete programmable framework to integrate different services and apis using protocols by providing similar user experience as zapier, "yahoo pipes" and "IFTTT". The main objective at the innovation perspective is to provide a graphical user interface for building data and fintech services mashups that aggregate open banking APIs, open available data sets and rules and creating Web based apps from various sources, and publishing those apps. The application worked by enabling users to "pipe" information from different open banking APIs and data sources and then set up rules for how that content should be modified.	Pilot 3, Any Pilot wants to provide Open Banking functionality	Crowdpolicy Open (Innovation) Banking Platform is based on the PSD2 directive API specification, providing API that can be integrated with any TPP that will be interested to utilize the open banking functionality for Account Information or Payment Initiaion Services.
<b>FBK</b>							
<b>a)</b>	AI-Engine-for-Psychometric-	7->8	Technology for extracting customer/consumer characteristics from their information in social media (e.g., Facebook Profile Photos). It will be				

D2.5 – Specifications of INFINITECH Technologies - I

	Profiling and Personalization		used to extract customer characteristics for services personalization (SIA & NBG Led Pilots).				
<b>b)</b>	Blockchain-based-Personal Data Market Infrastructure and Applications	7->7	Technology for executing encrypted queries over personal data in blockchain, while at the same time enabling a personal data market. It will be used in the blockchain pilots.				
<b>FTS</b>							
	OpenSource AI/ML frameworks	7->7	These frameworks facilitate the development of AI/ML based tools, which shall be applied to Financial Crime				
<b>GFT</b>							
<b>a)</b>	Data Layer - REST API	4	A Data Layer to support Security Data Model with REST API based on a not relational database (MongoDB). Supports heterogeneous sources. Developed upon FLASK-Python3 framework and dockerized to be deployed on Kubernetes infrastructure.				
<b>b)</b>	Front - End	6	Advanced Front End for Big Data management and processing				
<b>c)</b>	Antifraud system	5	Antifraud DB, Investigation tools, machine learning and scoring engine for real time prevention				
<b>GLA</b>							
	Terrier Information Retrieval Platform	5->7	Search Engine for BigData sets that offers integration with Spark for distributed processing	<a href="http://terrier.org">http://terrier.org</a>	We plan to extend Terrier with a new open source module that combines real-time data stream ingestion (via Apache Flink) with distributed database access (via LeanXcale) for real-time data indexing and updating from		

D2.5 – Specifications of INFINITECH Technologies - I

					multiple sources, within WP3 and WP5. We will also expand Terrier with enhanced Python integration, allowing easier use from common data science pipelines, such as those involving Pandas. This technology can be used for tasks such as searching/sampling financial product portfolios, user profiles or for providing recommendations.		
<b>GRAD</b>							
<b>a)</b>	Anonymization tool	5->6-7	A tool that anonymizes data in order to preserve privacy. It also provides metrics that allow to measure the risk of the anonymized data and the impact of the anonymization process on the utility of the data. The tool will be used in pilot #11 and pilot #12. The component needs a specific configuration/development for each pilot in which it is used	There is no public documentation available	In T2.3 the REST API of the component will be defined, so it can be instantiated within INFINITECH's platform. The rest of the enhancements of the tool will be related to developing new anonymization algorithms and metrics, which fall under the scope of T3.5.	Pilot #11 Pilot #12	The tool has a REST API, but it should be updated/reviewed
<b>LXS</b>							
<b>a)</b>	Polyglot Database Management System	4->7	The LXS DBMS is a polystore database that provides access to different and heterogeneous datastores via a common interface. It allows for the data user to submit a query, whose scan operators can request data that are stored in external datastores, and combine their intermediate results with data coming from other sources, either LXS internal datastore or others. For instance, a JOIN operator might require to JOIN table A (resigned in LXS) and table B (resigned in a MongoDB or a Hadoop DataLake). At this phase, there is a support for a limited target datastores, for Proof-of-Concept of the prototype. Moreover, the user has to write queries for the target datastores in the specific dialect. What is more, JOIN operations are not efficient, as they require all data resigned in an	There is no public documentation available	The enchantments that are planned to be implemented are the following: 1) Provide support for a variety of different datastores, according to the needs of the pilots. 2) introduce a novel SQL-like query language, so that the data user can write a query in a seamless way, and let the polystore interpret it to the target datastore. 3) improve the query engine in order to take into account scan operations (mostly part of the one of the JOIN arguments) that need to retrieve data from an external datastore. This will require the query		

D2.5 – Specifications of INFINITECH Technologies - I

			external datastore to be retrieved in the query engine level.		optimizer to explore equivalent operation graphs based on the nature of the target polystore, the operator to be able to push down operations to the target store, and the query processor to set up the corresponding data pipelines during the execution of the query plan. It is not clear at this moment, if the corresponding wrappers/connectors for each target datastore will be able to retrieve statistics to feed the query optimizer. This depends on the target datastores that will be used by the pilots, and their capabilities to expose this type of information		
<b>b)</b>	HTAP Database for the financial and insurance sector	4-6- >7-8	An ultra-scalable SQL Database and real-time big data platform that revolutionize the business database management systems by introducing the next generation business database that can scale in any of the three Vs of Big Data (Volume, Velocity and Variety). In more technical details, it provides an ultra-scalable transactional management system that can scale out to 100s of nodes, which is typically a bottleneck in traditional database systems that provide transactional semantics, while on the same time, is full SQL compatible and ensures all ACID transactional properties. It additionally exposes an interface for direct access of its key-value storage engine, thus providing a dual access without downgrade transactional semantics. It offers OLTP and OLAP integration, thus providing support for HTAP that allows for analytical queries over operational data, which realizes the concept of real time business intelligence. Finally, it enables for the execution of polyglot query processing across different and heterogeneous data sources. Modules of the	There is no public documentation available	LXS background technology will be enhanced in order to support the data management building blocks: Mainly, it will be enhanced in order to comply with the requirements for HTAP support, and to be compliant with the target data sources that need to be accessed via the polyglot mechanism. Moreover, it will provide support for real time query processing, enabling queries that combine both streaming data with data at rest. Finally, it will provide the support for incremental and parallel query analytics. However, at this phase of the project, it is not yet defined the exact technical details regarding these enhancements.		



D2.5 – Specifications of INFINITECH Technologies - I

			database will be used for implemented all building blocks of INFINITECH that are related to data management.				
<b>RB</b>							
<b>a)</b>	Natural Language Processing for real-time, high-accuracy Credit Risk assessment	3->8	ReportBrain's NLP functionalities: -Provide real-time structured feeds on risk-assessment worthy information sourced from the news -Interlink entities, updates and maintains knowledge graphs in real-time with all the interlinking of entities -Use “visual” algorithms to collect and analyse the news in real-time in 65 languages 24/7 -Classify articles in real-time by their content (politics, business etc.) -Use its own models, identifies entities (organizations, persons and locations) in real-time	Public Documentation under development	By using the AI enhancements that will be developed by Reportbrain in financial services & insurance sectors, users will be able to add an extra, yet orthocanonical feed to their existing credit rating models that will provide a real-time understanding of the world and more specifically on what’s if happening with specific entities (organizations & persons) of interest.		
<b>SILO</b>							
<b>a)</b>	Machine learning algorithms for health related data	6->7	SILO has implemented such approaches in different health-related projects such as CrowdHEALT	There is no public documentation available	SILO to make suggestions for the enhancement of the Platform	Pilot#12	
<b>UNIC</b>							
	Please state if you have some background technology?		If yes please provide info.		Please state on which area / technology / platform you would like to contribute to as part of your participation to the T2.3 specifications		
<b>WEA</b>							

## D2.5 – Specifications of INFINITECH Technologies - I

	Wenalyze Big data analytics platform		Platform that collect and process information from multiple open data sources regarding SMEs and apply cognitive algorithm to detect risk and changes in financial needs. The tool will be use in pilot 13	There is no public documentation available	As far as I understand, there won't be any relevant enhancements of the tool in T2.3.	#13	YES API-Rest
--	---	--	--	--	---	-----	--------------

## Background technology

Table 2 – Background technology

Company	Name of Tool / Platform	TRL level	Characteristics / Technology	Reference to Documentation / Links / Demo environment, other information ....	Ideas for enhancements (please make your suggestions)	Targeted pilots (pilots tools / platform aim at)	Availability of interfaces / API
<b>AGRO</b>							
a)	Octopush geospatial enabling framework	9	Octopush is a geospatial enabling framework, developed by AgroApps, allowing the collection, pre-processing, post-processing and distribution of geospatial data products and services, either referred to remote sensing (satellite, drones) acquisitions or multidimensional data outputs from numerical simulations. Octopush allows users to have access through a centralized access point to decentralized services, while Octopush SDK enables IT developers, to easily adapt or expand the provided geospatial services. Octopush was created by AgroApps aiming to address the company/operations and services need for a modular system, independent from any third-party service provider (excluding those offering raw data like Copernicus, NASA etc.), a framework that will be easily adapted to the market needs and follow the service-oriented business model of the company. Octopush is the baseline framework that addresses the Agri companies need in geospatial information, either	<a href="https://api.draxis.gr/">https://api.draxis.gr/</a>	n/a	#14	YES, RESTFUL API

## D2.5 – Specifications of INFINITECH Technologies - I

			through the development of new services and data models or the adaptation of the existing ones. Some of the services that are currently offered by Octopush are: a significant number of Vegetation Indices derived from Optical and SAR imagery and crop specific biophysical parameters including leaf area index, chlorophyll content and above ground biomass; crop specific yield estimation; farm management information services such as irrigation scheduling and variable rate fertilization; weather-driven models of possible pests and diseases outbreaks; high resolution weather forecasts and specific agrometeorological parameters; crop damage assessment services.				
<b>b)</b>	AgroApps Weather Intelligence Engine	9	Weather Intelligence Engine, developed by AgroApps, is a numerical weather prediction and atmospheric data assimilation processing chain, based on the WRF numerical weather prediction model. Weather Intelligence Engine is producing operationally all the needed weather data products (Near-real-time, medium-range weather forecasting, subseasonal to seasonal forecasts) by AgroApps offered services	<a href="https://api.draxis.gr/">https://api.draxis.gr/</a>	Weather Intelligence Service could take advantage of the available INFINITECH HPC resources, and pilot test a hybrid ensemble data assimilation scheme in convective scales.	#14	YES, RESTFULL API
<b>BOUN</b>							
<b>a)</b>	Partitioned and Distributed Transaction Graphs	4->7	AI Engine executing DL (deep neural networks) algorithms over semantically annotated streams. It will be used in the pilot for customer centric analytics (BOI led pilot).	BOUN to provide links / docs that can be shared	BOUN to make suggestions for the enhancement of the Platform		
<b>ENG</b>							
<b>a)</b>	Micro-service based platform for composition,	6->7	A micro-service based platform for composition, deployment, optimisation, execution and monitoring of big data analytics workflows (covering ingestion, preparation, analysis and visualization). It is designed and developed on top of the most cutting-edge open source Big Data technologies and framework.	<a href="https://home.alidalab.it/">https://home.alidalab.it/</a>	ENG to make suggestions for the enhancement of the Platform	#10	RESTful API RESTful API for BDA service management e workflow execution available

D2.5 – Specifications of INFINITECH Technologies - I

	deployment and execution of BDA applications						
<b>b)</b>	Knowage - open source suite for business analytics	6->7	Knowage is a business analytics suite that combines traditional and big data sources into valuable and meaningful information. A full set of features, such as data federation, mash-up, data/text mining and advanced data visualization, give comprehensive support to rich and multi-source data analysis.	<a href="https://www.knowage-suite.com/site/home/">https://www.knowage-suite.com/site/home/</a>			
<b>IBM</b>							
<b>a)</b>	Hyperledger Fabric	9->9	The permissioned blockchain infrastructures to be used in the project as part of the INFINITECH data exchange technologies and in relevant pilots (like the SIA led pilot). It is open source and promoted by IBM and the Linux Foundation.	<a href="https://www.ibm.com/blockchain/hyperledger">https://www.ibm.com/blockchain/hyperledger</a>			
<b>ISPRINT</b>							
<b>a)</b>	Healthentia LifeSciences BigData Platform	6->7	BigData platform providing data sources aggregation and management, as well as tools for analytics and visualization. It will be used in the IoT-based Life Insurance pilot (ARI/RRD Lead pilot).	<a href="https://healthentia.com/">https://healthentia.com/</a>	Re-purpose the platform to support the insurance-related monitoring of people for pilot 12		
<b>JSI</b>							
<b>a)</b>	Event-registry	9	Event Registry (ER) is real-time cross-lingual global media monitoring service for modelling global social dynamics (eventregistry.org) developed by the JSI. ER aggregates and analyses news content for over 120,000 news sources published globally in 100+ languages. Events mentioned in the news are identified and relevant information about them is automatically extracted and stored in a searchable form. The data can be accessed directly on the platform or via the API. ER supports various analytics including deep	<a href="http://eventregistry.org/">http://eventregistry.org/</a>	Eventregistry could be enriched with new insights of the potential use scenarios and would benefit from them to expand the current offering to the fintech industry. In addition, better insight will be given to the team in order to develop further analytics.		

D2.5 – Specifications of INFINITECH Technologies - I

			analytics of the events and correlations between events, extracted entities and financial data extracted from the main financial indexes.				
<b>b)</b>	Stream Story	4	Stream-Story multi-resolution modelling and explanation of (possibly real-time) streaming data: (1)Exploratory data mining - A system for the analysis of multivariate time series. It computes and visualizes a hierarchical Markov chain model which captures the qualitative behaviour of the systems' dynamics.; (2) Multi-scale representation -The hierarchical model allows users to interactively find suitable scales for interpreting the data; (3) Real-time monitoring Visualizes streaming data by mapping it to the hierarchical model. It can provide predictions and alarms for different behaviour.	<a href="http://streamstory.js.si/">http://streamstory.js.si/</a>			
<b>c)</b>	Qminer	7->9	Bigdata				
<b>NUIG</b>							
<b>a)</b>	SSC - Super Stream Collider, a multiform at Data Management and Query System	6->7	The SSC enables distributed cloud- based high-performance processing of semantically linked streams i.e. it is an enabler for semantic analytics. It will be used for analytics over semantically unified/interoperable streams (in WP5), as well as in the KYC and customer-centric services pilot (BOI-led pilot).	NUIG to provide links / docs that can be shared	NUIG to make suggestions for the enhancement of the Platform		
<b>b)</b>	The Global Engine with Neural Network Intelligence (GENNI)	4->7	AI Engine executing DL (deep neural networks) algorithms over semantically annotated streams. It will be used in the pilot for customer centric analytics (BOI led pilot).	NUIG to provide links / docs that can be shared	NUIG to make suggestions for the enhancement of the Platform		
<b>UBI</b>							

D2.5 – Specifications of INFINITECH Technologies - I

<p>a)</p>	<p>Data Check-in mechanism</p>	<p>A sophisticated data check-in mechanism that is enabling the preparation and uploading of the data provider’s (public or confidential) datasets in the cloud platform that is one of the results of the ICARUS H2020 project. The data check-in mechanism is deployed on the premises of the data provider as a stand-alone desktop application and receives as input a list of data check-in jobs that incorporate a set of instructions with all the actions that will be performed on a specific dataset, residing on the local storage of running operating system, in order to enable the data preparation and uploading of new datasets in a secure manner. Internally, the mechanism handles the orchestration and execution of the designed instructions with the use of incorporated (micro) services for the: a) data mapping of data source entities to the designed common data schema, b) data cleaning operations on the data source entities, c) the anonymisation operations on the data source entities and d) the encryption of the data source entities . This list of (micro) services is expandable based on the needs of each platform. The data check-in mechanism is offered in the form of a local client for all OS (Mac, Linux, Windows) and is designed and developed using the latest technologies for desktop apps with the aim to offer end-to-end security on the data preparation and data upload tasks.</p>	<p>There is no public documentation available at the moment.</p>	<p>At this point enhancements cannot be proposed. The enhancements of this tool will depend on the outcomes of the reference architecture and the needs that will arise for the data check-in process.</p>	<p>Currently only under the scope of Pilot 6</p>	<p>The data check-in mechanism is offered in the form of a local client for all OS (Mac, Linux, Windows). The client is offering a list of APIs that are secured with JWT technology and are utilised in order to communicate with client for: a) the submission of the instructions that will be performed locally for the data preparation and b) the active monitor and reporting of the performed activities based on the provided instructions. Furthermore, upon the proper configuration the client is able to communicate with the preconfigured system in order to upload the produced datasets into the system</p>
-----------	--------------------------------	--	--	--	--	--

D2.5 – Specifications of INFINITECH Technologies - I

<b>UNP</b>							
<b>a)</b>	IoT-Catalogue	9	The 'IoT Catalogue' is a web-based catalogue from where to pick & choose IoT solutions; it is an explorer of innovations in IoT applications and technologies. It aims to be single entry point of support to IoT developers/integrators/advisors/end-users in the process of identifying and selecting IoT technologies (ranging from complete end-to-end solutions to tools and components/parts) but also inspecting a wide set of IoT use-cases, their validations, associated contact persons/organisations, detailed characterisation (value propositions, ICT problems, functions, target, domain), supporting technological solutions, and much more.	<a href="http://www.iot-catalogue.com">www.iot-catalogue.com</a>			
<b>UPRC</b>							
<b>a)</b>	Analytics Library	4 -> 5	Participating in ATMOSPHERE (Adaptive, Trustworthy, Manageable, Orchestrated, Secure Privacy-assuring Hybrid, Ecosystem for RESilient Cloud Computing) EU project, the UPRC team, among others, focused on the delivery of the library of services, which can be utilized as a baseline for the INFINITECH library. (WP5)	<a href="https://www.atmosphere-eubrazil.eu/">https://www.atmosphere-eubrazil.eu/</a>	Update the library to include metadata relevant to security and privacy constraints of the INFINITECHJ algorithms to be made available through the library		
<b>b)</b>	Catalogue of Objects	5 -> 6	During the 5GTANGO and MATILDA EU Projects, both aiming to enable the flexible programmability of 5G networks and to devise and realize a radical shift in the development of software for 5G-ready applications, UPRC team contributed to the market platform and to the catalogue of services & functions, respectively. Therefore, the relative outcomes and the solutions developed by UPRC for these projects will be utilized as a baseline for the INFINITECH's marketplace, integrating with UNP's IoT-catalogue. (WP8)	<a href="https://www.5gtango.eu/">https://www.5gtango.eu/</a> <a href="https://www.matilda-5g.eu/">https://www.matilda-5g.eu/</a>	Extend to support a variety of assets (e.g. datasets, models, etc) as well as to support composite assets (i.e. analytics pipelines)		

### 3. INFINITECH components specifications

The Infinitech Reference Architecture (IRA) [Deliverable D2.13], defines a set of layers and allows flexible **workflows** out of data processing modules, called **components** that performs specific data transformations. A workflow usually consumes one set of data (sources) to produce another set of data (destinations). A component can be identified by the Input/Output interfaces and the functionality (transformation) provided at the edge. In that respect, a data component can be considered as a black box that can be replaced by any technological implementation that performs at the edge the same transformation.

The figure below depicts the different reference layers of the IRA along with examples of data components. Scope of the section is to identify existing and desired data components that will support the design, the development and deployment of the Pilots’ sandboxes.

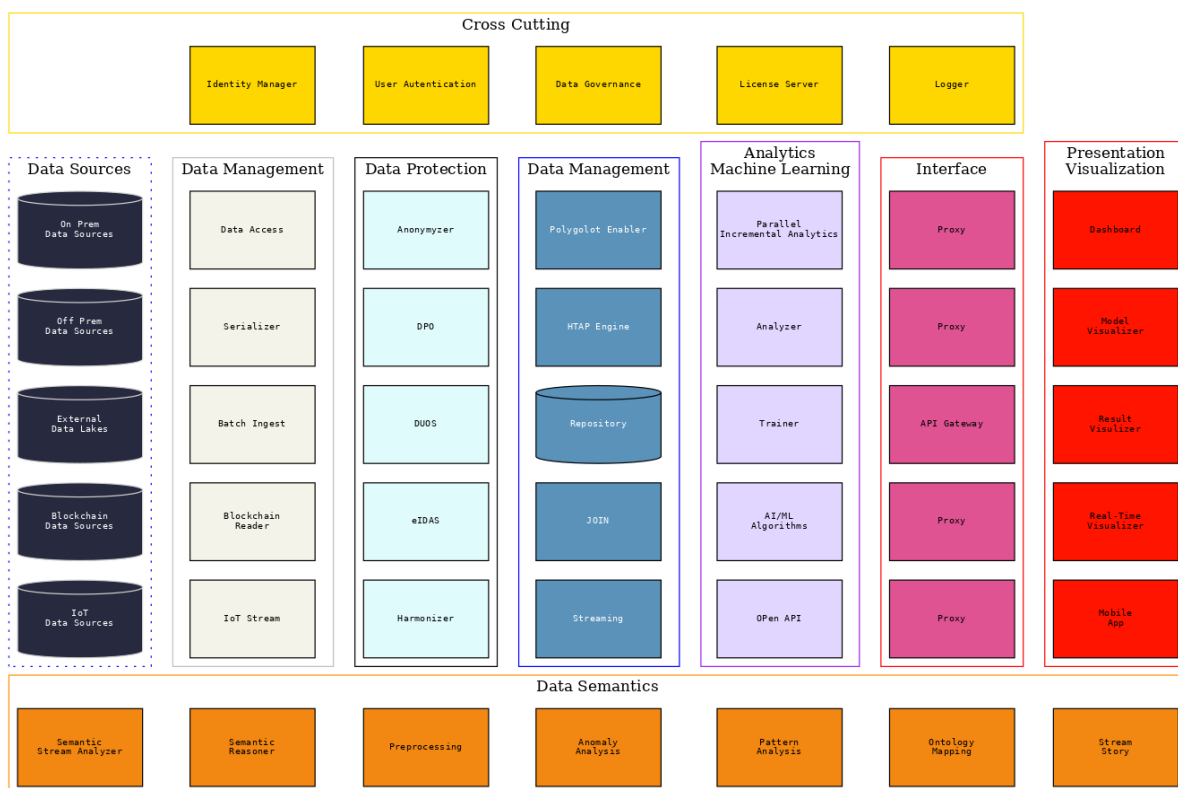


Figure 2 – Logical Schema of Data Processing Components in INFINITECH RA

The “black-boxes” follow predefined colors in order to map components in the different layers.



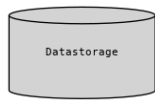
An INFINITECH solution (sandbox) can be built organizing components in a workflow (sometimes referred to data pipeline) to accomplish a complex transformation from one set of data sources to another set of data solving the business case. Components should therefore be interoperable, with clear interfaces and perform a clear function over the data. The rest of this section is a first attempt to identify the existing and to be developed components of the project. Further progress in the project will provide the needed details for implementation and deployment.



## Template for Data Component Description

In the following a first list of the identified components, are described using a standard template described in the table below.

Table 3 - Component Description Template

Attribute	Documentation & Example
<b>Component ID</b>	INF-DCO-001
<b>Component Name</b>	NAME
<b>Description</b>	Short description
<b>Icon</b>	 (non mandatory)  , 
<b>IP Owner &amp; Partner in Charge</b>	
<b>INFINITECH Component Category</b>	
<b>IRA - BDVA Layer</b>	Infrastructure/Data Management/Data Processing
<b>Input (Required by the Component)</b>	Description of Input format (File, JSON, ...)
<b>Output (Produced by the Component)</b>	Description of Output format (File, JSON, ...)
<b>Technology or Platform to be used</b>	
<b>Part of INFINITECH Core</b>	Yes / No
<b>License</b>	Opensource License or Closed
<b>Other Information / Remarks</b>	E.g. Project's official website
<b>Detailed Documentation</b>	Link to reference documentation if any

## 3.1 Component Group: Data Management

### 3.1.1 Relational Database

Attribute	Documentation & Example
<b>Component ID</b>	INF-DMA-108
<b>Component Name</b>	Relational Database
<b>Description</b>	This component consists in the central data repository of the platform. It enables transactional semantics and provides capabilities for query processing based on standard SQL statements. It can scale out on the runtime while continuing serving operational workloads. It can support analytical processing in combination with operational data modifications with the level of isolation to be <i>snapshot isolation</i> . That is, it enables real-time business analytics.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LXS
<b>INFINITECH Component Category</b>	Data Management
<b>IRA - BDVA Layer</b>	Data Processing
<b>Input (Required by the Component)</b>	A JDBC connection needs to be established in order to execute statements for storing the data.
<b>Output (Produced by the Component)</b>	A JDBC ResultSet or a key-value JSON, based on the OData specification.
<b>Technology or Platform to be used</b>	The data repository is a written in Java and C, and provides support for JDBC, ODBC and python drivers. It runs on K8S cluster.
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	This component is under proprietary rights of LXS.
<b>Other Information / Remarks</b>	
<b>Detailed Documentation</b>	N/A

### 3.1.2 Polyglot Query Processing

Attribute	Documentation & Example
<b>Component ID</b>	INF- DMA -109
<b>Component Name</b>	Polyglot Query Processing
<b>Description</b>	This component enables the query execution over more than one datastore in seamless manner. The data user can submit a single statement and let this component to execute the query by pushing it down to the target database. By doing this, it can process the data on

	premise, and retrieve only the results, thus it is convenient for cases where data cannot be loaded the platform and they need to be accessed from an external datastore.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LXS
<b>INFINITECH Component Category</b>	Data Management
<b>IRA - BDVA Layer</b>	Data Processing
<b>Input (Required by the Component)</b>	This component needs to open a connection with the target database. It will require different input according to the number of target stores that will be supported.
<b>Output (Produced by the Component)</b>	A JDBC ResultSet or a key-value JSON, based on the OData specification.
<b>Technology or Platform to be used</b>	This component is implemented in Java. It is an extension of the central repository itself. In fact, it consists of a jar binary which is loaded in the classpath of the query engine of the data repository. As of that, it will be deployed as part of the query engine via K8S cluster.
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	This component is under proprietary rights of LXS.
<b>Other Information / Remarks</b>	
<b>Detailed Documentation</b>	N/A

### 3.1.3 Incremental Analytics

<b>Attribute</b>	<b>Documentation &amp; Example</b>
<b>Component ID</b>	INF- DMA -110
<b>Component Name</b>	Incremental Analytics
<b>Description</b>	This component enables the query execution in an incremental fashion. The data user will be able to submit a continuous query to the datastore, which will be continuously and incrementally validated. This means that the initial results will be retrieved first, and as data arrives to the data repository, they will be validated against the submitted query, and if it validates the statement, it will be returned to the user.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LXS
<b>INFINITECH Component Category</b>	Data Management
<b>IRA - BDVA Layer</b>	Data Processing

<b>Input (Required by the Component)</b>	A JDBC connection needs to be established in order to execute statements for storing the data.
<b>Output (Produced by the Component)</b>	To be defined
<b>Technology or Platform to be used</b>	This component is implemented in Java. It is an extension of the central repository itself. In fact, it consists of a jar binary which is loaded in the classpath of the query engine of the data repository. As of that, it will be deployed as part of the query engine via K8S cluster.
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	This component is under proprietary rights of LXS.
<b>Other Information / Remarks</b>	
<b>Detailed Documentation</b>	N/A

### 3.1.4 Filtering

<b>Attribute</b>	<b>Documentation &amp; Example</b>
<b>Component ID</b>	INF- DMA -111
<b>Component Name</b>	Filtering
<b>Description</b>	The purpose of this component is to be able to select specific rows from a datatable according to some filters. Filters should be understood as a logical operation (or a combination of logical operations) where each operand will be a value or the name of a column.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LXS
<b>INFINITECH Component Category</b>	Data Management
<b>IRA - BDVA Layer</b>	Data Processing
<b>Input (Required by the Component)</b>	A JDBC connection needs to be established in order to execute statements for storing the data.
<b>Output (Produced by the Component)</b>	A JDBC ResultSet or a key-value JSON, based on the OData specification.
<b>Technology or Platform to be used</b>	This component is implemented in Java and it is part of the query engine of the data repository itself. The query engine is based on the Apache Calcite and has been extended in order to provide full SQL query support, along with an implementation of the data adapter to push down the corresponding query statements to the data nodes. It will be deployed as part of the query engine via K8S cluster.
<b>Part of INFINITECH Core</b>	Yes

<b>License</b>	This component is under proprietary rights of LXS.
<b>Other Information / Remarks</b>	
<b>Detailed Documentation</b>	N/A

### 3.1.5 Joiner

Attribute	Documentation & Example
<b>Component ID</b>	INF- DMA -112
<b>Component Name</b>	Joiner
<b>Description</b>	The purpose of this component is to be able to execute the JOIN operation between two tables. Each of the two operators of this operation can be a data table of the data repository itself, or of an external data store. Its internal query optimizer can decide on the optimal execution plan to follow and apply different implementations like <i>nested loop joins</i> , <i>merge joins</i> , <i>bind joins</i> etc according to the statistics and available indexing. It can also rewrite the original plan to decide on an equivalent query execution plan. This will depend on the statistical information that is available for each table, and the amount of supported query processing, that can allow optimizations by pushing down operators to the storage.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LXS
<b>INFINITECH Component Category</b>	Data Management
<b>IRA - BDVA Layer</b>	Data Processing
<b>Input (Required by the Component)</b>	A JDBC connection needs to be established in order to execute statements for storing the data.
<b>Output (Produced by the Component)</b>	A JDBC ResultSet or a key-value JSON, based on the OData specification.
<b>Technology or Platform to be used</b>	This component is implemented in Java and it is part of the query engine of the data repository itself. The query engine is based on the Apache Calcite and has been extended in order to provide full SQL query support, along with an implementation of the data adapter to push down the corresponding query statements to the data nodes. It will be deployed as part of the query engine via K8S cluster.
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	This component is under proprietary rights of LXS.
<b>Other Information / Remarks</b>	
<b>Detailed Documentation</b>	N/A

## 3.1.6 OneHotEncoder

Attribute	Documentation & Example
<b>Component ID</b>	INF- DMA -113
<b>Component Name</b>	OneHotEncoder
<b>Description</b>	Service to represent categorical variables as binary vectors
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ENG
<b>INFINITECH Component Category</b>	Data Management
<b>IRA - BDVA Layer</b>	Processing
<b>Input (Required by the Component)</b>	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE
<b>Output (Produced by the Component)</b>	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE
<b>Technology or Platform to be used</b>	It is a BDA service registered in the ALIDA catalogue. It is a PySpark-based micro-service running on K8S cluster mode.
<b>Part of INFINITECH Core</b>	This component is general purpose and reusable as a BDA service within ALIDA
<b>License</b>	Component developed with open source technologies. ENG Proprietary.
<b>Other Information / Remarks</b>	<p>In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies.</p> <p>This service is registered in ALIDA catalogue as Spring Boot Application containing the python code and its dependencies. After implementing the algorithm using Pyspark, creating the Dockerfile and pushing the new image inside a repository, this microservice is registered into the ALIDA catalogue through the GUI.</p> <p>Source: <a href="https://home.alidalab.it/">https://home.alidalab.it/</a></p>
<b>Detailed Documentation</b>	N/A

## 3.1.7 Visualization Preparation

Attribute	Documentation & Example
<b>Component ID</b>	INF- DMA -114
<b>Component Name</b>	Visualization Preparation
<b>Description</b>	Stream BDA Service: Service to prepare data to the visualization. It will be able to choose between different preparation algorithms depending on the type of incoming data or the data you want to view
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ENG
<b>INFINITECH Component Category</b>	Data Management
<b>IRA - BDVA Layer</b>	Processing
<b>Input (Required by the Component)</b>	This component is a consumer who subscribes to Kafka's topic, so it reads the messages that will be sent to the topic.
<b>Output (Produced by the Component)</b>	This component is also a producer, who write the results to other one Kafka topic.
<b>Technology or Platform to be used</b>	It will be a microservice belonging to the ALIDA core.
<b>Part of INFINITECH Core</b>	This component is part of ALIDA core
<b>License</b>	Component developed with open source technologies. ENG Proprietary
<b>Other Information / Remarks</b>	In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies. Source: <a href="https://home.alidalab.it/">https://home.alidalab.it/</a>
<b>Detailed Documentation</b>	N/A

### 3.1.8 Stream Processor

Attribute	Documentation & Example
<b>Component ID</b>	INF- DMA -115
<b>Component Name</b>	Stream Processor
<b>Description</b>	This component will provide streaming processing capabilities. The data user can declare continuous queries that will be executed over the data stream. It will also allow to combine streaming data with data <i>at-rest</i> and


	also enable the storage of data streams even when injected in very high rates.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LXS
<b>INFINITECH Component Category</b>	Data Management
<b>IRA - BDVA Layer</b>	Data Processing
<b>Input (Required by the Component)</b>	This component receives data from a stream and can consume data stored at rest in the data repository
<b>Output (Produced by the Component)</b>	Depends on the pilot
<b>Technology or Platform to be used</b>	This component will be based on Apache Flink and will be containerized in order to be deploy with a K8S cluster.
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	This component is under proprietary rights of LXS.
<b>Other Information / Remarks</b>	This component can exploit the capabilities of the declarative real time analytics. This will be very useful in cases the data user wants to calculate a value over a stream that will need to perform an expensive scan operation over a data table (i.e. compare the input with the overall average of a field in a table). As scan operations (and operations that will require a scan operation, as the average etc) have a complexity bigger then $O(1)$ , they are timely costly and cannot be executed in a stream. For that, the developer often caches that value and periodically updates the value. With the declarative real-time analytics, the data user is given the opportunity to declare such an analytical operation (i.e. the overall average) with an SQL fashion, and the query will be executed with a complexity of a get operation, which will allow these types of analytical operations to be included in a stream operation, providing to it the current average, with respect to data consistency and isolation in terms of ACID properties and transactional semantics.
<b>Detailed Documentation</b>	N/A

## 3.2 Component Group: Security and Privacy

### 3.2.1 eIDAS integration

Attribute	Documentation & Example
<b>Component ID</b>	INF-DPR-116
<b>Component Name</b>	SPeIDI (Service Provider eIDAS Integration)
<b>Description</b>	SPeIDI is a solution for authenticating against the eIDAS infrastructure providing strong cross-border authentication



	<p>It provides Authentication for citizens trying to access other country services different of their origin country</p> <p>It is an Adapter that eases Service Providers connection to eIDAS Network for authentication purposes.</p> <p>It allows the use of eID issued by European National authorities according to the EU eID schemas.</p> <p>Creates and sends SAML Request to eIDAS node, validates and translate SAML Response from SAML 2.0 to JSON and forward it to the relevant SP service</p> <p>SP infrastructure independent e.g. Java, Non-Java...</p> <p>Provides UI easing usability and privacy: User country selection, displays list of mandatory and optional attributes requested by the SP</p> <p>Asks for user consent, provides privacy policy from the SP</p>
<b>Icon</b>	<p><b>SPeIDI</b></p> 
<b>IP Owner &amp; Partner in Charge</b>	Atos
<b>INFINITECH Component Category</b>	Data Protection
<b>BDVA Layer</b>	Infrastructure
<b>Input (Required by the Component)</b>	JWT (Java Web Token) based Request from SP. Possible customization to other communications with SPs
<b>Output (Produced by the Component)</b>	JWT (Java Web Token) based Response to SP. . Possible customization to other communications with SPs.
<b>Technology or Platform to be used</b>	eIDAS infrastructure and Service Provider technologies (integration with it)
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	EUPL
<b>Other Information / Remarks</b>	
<b>Detailed Documentation</b>	


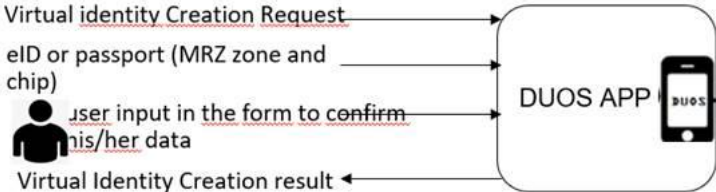
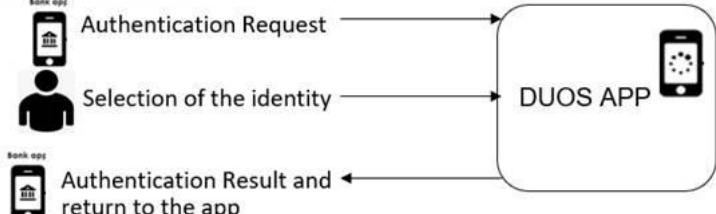
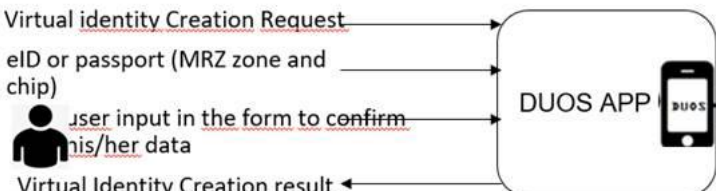
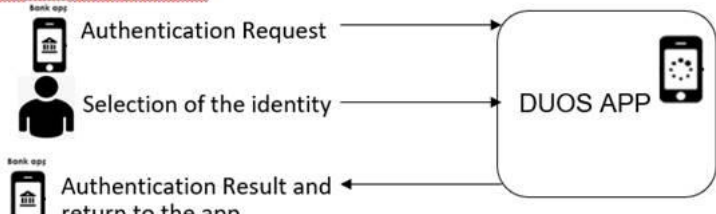
### 3.2.2 Data Anonymization

<b>Attribute</b>	<b>Documentation &amp; Example</b>
<b>Component ID</b>	INF- DPR -117
<b>Component Name</b>	Anonymization tool
<b>Description</b>	The anonymization tool modifies data in order to preserve privacy. It is especially indicated in those cases where a dataset contains personal data and it has to be outsourced or shared with a third party. The tool includes different anonymization

	algorithms that aim at avoiding the appearances of data combinations that could lead to a possible re-identification of the data subjects. It also includes a set of privacy and utility metrics that allow to measure the risk that remains after anonymizing the dataset, and the impact of the anonymization process on the quality of the data.
<b>IP Owner &amp; Partner in Charge</b>	GRAD
<b>INFINITECH Component Category</b>	Data Protection
<b>IRA - BDVA Layer</b>	Data
<b>Input (Required by the Component)</b>	The component requires two inputs: the data that has to be anonymized and a configuration file that defines the structure of the data, its location and the privacy requirements.
<b>Output (Produced by the Component)</b>	It stores the anonymized data in a database. The location of the database has to be known beforehand (through the configuration file that is taken as an input).
<b>Technology or Platform to be used</b>	The tool is based on two modules, developed in Java and Python. The Java module can be deployed as a Docker container. The tool requires to retrieve and store the raw and anonymized version of the data from a relational database.
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	Closed
<b>Other Information / Remarks</b>	Even though the tool will be part of the INFINITECH platform, it will not be available for all the pilots, as it needs a specific configuration (and probably development) for each specific use case.
<b>Detailed Documentation</b>	N/A

### 3.2.3 Digital User Onboarding System (DUOS)

Attribute	Documentation & Example
<b>Component ID</b>	INF- DPR -118
<b>Component Name</b>	DUOS (Digital User On-boarding Services)
<b>Description</b>	<p>Provides remote user registration using eID or passport, dealing with virtual identities in a mobile device</p> <p>It uses various identity proofing and verification services that link new user eID creation (virtual or derived eID) with government issued e-ID</p> <p>It verifies electronic data stored on chip and machine readable zone</p> <p>Provides Flexible Multi-factor authentication for different users or identities</p>


<p><b>Icon</b></p>	
<p><b>IP Owner &amp; Partner in Charge</b></p>	<p>Atos</p>
<p><b>INFINITECH Component Category</b></p>	<p>Data Protection</p>
<p><b>BDVA Layer</b></p>	<p>Infrastructure</p>
<p><b>Input (Required by the Component)</b></p>	<p><b>Virtual Identity Creation</b></p> <p>Virtual identity Creation Request →</p> <p>eID or passport (MRZ zone and chip) →</p> <p>User input in the form to confirm his/her data →</p> <p>Virtual Identity Creation result ←</p>  <p><b>Authentication</b></p> <p>Bank app Authentication Request →</p> <p>User Selection of the identity →</p> <p>Bank app Authentication Result and return to the app ←</p> 
<p><b>Output (Produced by the Component)</b></p>	<p><b>Virtual Identity Creation</b></p> <p>Virtual identity Creation Request →</p> <p>eID or passport (MRZ zone and chip) →</p> <p>User input in the form to confirm his/her data →</p> <p>Virtual Identity Creation result ←</p>  <p><b>Authentication</b></p> <p>Bank app Authentication Request →</p> <p>User Selection of the identity →</p> <p>Bank app Authentication Result and return to the app ←</p> 
<p><b>Technology or Platform to be used</b></p>	<p>Android</p>
<p><b>Part of INFINITECH Core</b></p>	<p>Yes</p>
<p><b>License</b></p>	<p>Different underlying licenses (Apache 2.0 license for MRZ reader and license to be decided for biometric checking)</p>
<p><b>Other Information / Remarks</b></p>	
<p><b>Detailed Documentation</b></p>	

### 3.2.4 Data Protection Orchestrator (DPO)

Attribute	Documentation & Example
<b>Component ID</b>	INF- DPR -119
<b>Component Name</b>	DPO (Data Protection Orchestrator)
<b>Description</b>	The DPO embeds and automats the assurance of security and privacy by design and by default in complex business flows. It provides orchestration of Privacy Enhancing Technologies and related services using BPM tools in order to integrate privacy or data protection perspectives into business processes. It can orchestrate any kind of REST service.
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	Atos
<b>INFINITECH Component Category</b>	Data Protection
<b>BDVA Layer</b>	Infrastructure
<b>Input (Required by the Component)</b>	REST call to the DPO and input data in JSON
<b>Output (Produced by the Component)</b>	Result of the execution in JSON
<b>Technology or Platform to be used</b>	The DPO interacts with the privacy expert that prepares the business flow in a BPMN file. The flow interacts with PETs such as anonymization
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	Apache 2.0
<b>Other Information / Remarks</b>	
<b>Detailed Documentation</b>	

## 3.3 Component Group: Blockchain and Information Sharing


### 3.3.1 Blockchain Reader

Attribute	Documentation & Example
<b>Component ID</b>	INF-BKC-120
<b>Component Name</b>	Blockchain Reader
<b>Description</b>	Fetches requested data from blockchain ledger.
<b>Icon</b>	

<b>IP Owner &amp; Partner in Charge</b>	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
<b>INFINITECH Component Category</b>	Blockchain
<b>BDVA Layer</b>	Infrastructure/Data
<b>Input (Required by the Component)</b>	A JSON file <ul style="list-style-type: none"> <li>the requested data details to fetch (query).</li> </ul>
<b>Output (Produced by the Component)</b>	A JSON file <ul style="list-style-type: none"> <li>Including any requested block information (stored in the ledger) that can be retrieved from the blockchain network and</li> <li>the transactions hashes, to verify that they were successfully validated by the blockchain network and stored on the corresponding ledger</li> </ul>
<b>Technology or Platform to be used</b>	Hyperledger
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	Apache 2.0
<b>Other Information / Remarks</b>	<ul style="list-style-type: none"> <li>Project's official website: <a href="https://www.hyperledger.org/">https://www.hyperledger.org/</a></li> <li>License: <a href="https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)">https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)</a></li> </ul>
<b>Detailed Documentation</b>	<ul style="list-style-type: none"> <li><a href="https://hyperledger.github.io/">https://hyperledger.github.io/</a></li> <li><a href="https://github.com/hyperledger/fabric">https://github.com/hyperledger/fabric</a></li> <li><a href="https://wiki.hyperledger.org/display/fabric">https://wiki.hyperledger.org/display/fabric</a></li> <li><a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html</a></li> <li><a href="https://www.hyperledger.org/projects/fabric">https://www.hyperledger.org/projects/fabric</a></li> <li><a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html</a></li> <li><a href="https://github.com/hyperledger/blockchain-explorer">https://github.com/hyperledger/blockchain-explorer</a></li> <li><a href="https://wiki.hyperledger.org/display/explorer">https://wiki.hyperledger.org/display/explorer</a></li> <li>Documentation of Input and Output Parameters: <a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html</a></li> </ul>


### 3.3.2 Blockchain Writer

Attribute	Documentation & Example
<b>Component ID</b>	INF- BKC -121
<b>Component Name</b>	Blockchain Writer
<b>Description</b>	Submits transactions on the blockchain ledger.


Icon	
IP Owner & Partner in Charge	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
INFINITECH Component Category	Blockchain
BDVA Layer	Infrastructure/Data
Input (Required by the Component)	A JSON file <ul style="list-style-type: none"> <li>Including the necessary data for the transaction submission.</li> </ul>
Output (Produced by the Component)	A JSON file <ul style="list-style-type: none"> <li>Including the transactions hashes, to verify that they were successfully validated by the blockchain network and stored on the corresponding ledger,</li> <li>(optional) the submitted data just for validation of correct input (the just now stored data in the ledger after it was approved by the consensus mechanism)</li> </ul>
Technology or Platform to be used	Hyperledger
Part of INFINITECH Core	Yes
License	Apache 2.0
Other Information / Remarks	<ul style="list-style-type: none"> <li>Project's official website: <a href="https://www.hyperledger.org/">https://www.hyperledger.org/</a></li> <li>License: <a href="https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)">https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)</a></li> </ul>
Detailed Documentation	<ul style="list-style-type: none"> <li><a href="https://hyperledger.github.io/">https://hyperledger.github.io/</a></li> <li><a href="https://github.com/hyperledger/fabric">https://github.com/hyperledger/fabric</a></li> <li><a href="https://wiki.hyperledger.org/display/fabric">https://wiki.hyperledger.org/display/fabric</a></li> <li><a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html</a></li> <li><a href="https://www.hyperledger.org/projects/fabric">https://www.hyperledger.org/projects/fabric</a></li> <li><a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html</a></li> <li><a href="https://github.com/hyperledger/blockchain-explorer">https://github.com/hyperledger/blockchain-explorer</a></li> <li><a href="https://wiki.hyperledger.org/display/explorer">https://wiki.hyperledger.org/display/explorer</a></li> <li>Documentation of Input and Output Parameters: <a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html</a></li> </ul>

### 3.3.3 Smart Contract Executor

Attribute	Documentation & Example
Component ID	INF- BKC -122
Component Name	Smart Contract Executor
Description	Executes smart contracts on the blockchain ledger.

<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
<b>INFINITECH Component Category</b>	Blockchain
<b>BDVA Layer</b>	Infrastructure/Data
<b>Input (Required by the Component)</b>	A JSON file <ul style="list-style-type: none"> <li>the chaincode or smart contract specific input required to submit transactions and deploy smart contracts.</li> </ul>
<b>Output (Produced by the Component)</b>	A JSON file <ul style="list-style-type: none"> <li>Including the smart contracts' transactions hashes, to verify that they were successfully deployed on the respective blockchain network,</li> <li>(optional) the submitted data just for validation of correct input (the just now stored data in the ledger after it was approved by the consensus mechanism)</li> </ul>
<b>Technology or Platform to be used</b>	Hyperledger
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	Apache 2.0
<b>Other Information / Remarks</b>	<ul style="list-style-type: none"> <li>Project's official website: <a href="https://www.hyperledger.org/">https://www.hyperledger.org/</a></li> <li>License: <a href="https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)">https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)</a></li> </ul>
<b>Detailed Documentation</b>	<ul style="list-style-type: none"> <li><a href="https://hyperledger.github.io/">https://hyperledger.github.io/</a></li> <li><a href="https://github.com/hyperledger/fabric">https://github.com/hyperledger/fabric</a></li> <li><a href="https://wiki.hyperledger.org/display/fabric">https://wiki.hyperledger.org/display/fabric</a></li> <li><a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html</a></li> <li><a href="https://www.hyperledger.org/projects/fabric">https://www.hyperledger.org/projects/fabric</a></li> <li><a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html</a></li> <li><a href="https://wiki.hyperledger.org/display/fabric">https://wiki.hyperledger.org/display/fabric</a></li> <li><a href="https://github.com/hyperledger/fabric-chaincode-node">https://github.com/hyperledger/fabric-chaincode-node</a></li> <li><a href="https://github.com/hyperledger/blockchain-explorer">https://github.com/hyperledger/blockchain-explorer</a></li> <li><a href="https://wiki.hyperledger.org/display/explorer">https://wiki.hyperledger.org/display/explorer</a></li> <li>Documentation of Input and Output Parameters: <a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html</a></li> </ul>


## 3.3.4 Blockchain Data Visualizer

Attribute	Documentation & Example
<b>Component ID</b>	INF- BKC -123
<b>Component Name</b>	Blockchain Data Visualizer
<b>Description</b>	Queries and displays information about blocks, transactions, chaincodes and transaction families, network name, status and nodes list, organizations list and peers list.
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
<b>INFINITECH Component Category</b>	Blockchain
<b>BDVA Layer</b>	Infrastructure/Data
<b>Input (Required by the Component)</b>	<p>A JSON file</p> <ul style="list-style-type: none"> <li>• Including the credentials pair (username &amp; password) that authenticate the user to have access to this component's "Output",</li> <li>• the blockchain network name and</li> <li>• the component's URL.</li> </ul>
<b>Output (Produced by the Component)</b>	<p>A web page</p> <ul style="list-style-type: none"> <li>• Including statistics of the blockchain network activity in charts and pies and</li> <li>• ability to view and query blocks, transactions, chaincodes and transaction families, network name, status and nodes list, organizations list and peers list.</li> </ul>
<b>Technology or Platform to be used</b>	Hyperledger
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	Apache 2.0
<b>Other Information / Remarks</b>	<ul style="list-style-type: none"> <li>• Project's official website: <a href="https://www.hyperledger.org/">https://www.hyperledger.org/</a></li> <li>• License: <a href="https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)">https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)</a></li> </ul>
<b>Detailed Documentation</b>	<ul style="list-style-type: none"> <li>• <a href="https://github.com/hyperledger/blockchain-explorer">https://github.com/hyperledger/blockchain-explorer</a></li> <li>• <a href="https://wiki.hyperledger.org/display/explorer">https://wiki.hyperledger.org/display/explorer</a></li> <li>• <a href="https://medium.com/@thanawitsupinpong/setting-up-hyperledger-explorer-on-fabric-5f1f7cda73b3">https://medium.com/@thanawitsupinpong/setting-up-hyperledger-explorer-on-fabric-5f1f7cda73b3</a></li> <li>• <a href="https://hyperledger.github.io/">https://hyperledger.github.io/</a></li> <li>• <a href="https://github.com/hyperledger/fabric">https://github.com/hyperledger/fabric</a></li> <li>• <a href="https://wiki.hyperledger.org/display/fabric">https://wiki.hyperledger.org/display/fabric</a></li> </ul>




	<ul style="list-style-type: none"> <li>• <a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html</a></li> <li>• <a href="https://www.hyperledger.org/projects/fabric">https://www.hyperledger.org/projects/fabric</a></li> <li>• <a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html</a></li> <li>• <a href="https://wiki.hyperledger.org/display/fabric">https://wiki.hyperledger.org/display/fabric</a></li> <li>• Documentation of Input and Output Parameters: <a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html</a></li> </ul>
--	---

### 3.3.5 Blockchain Authenticator

Attribute	Documentation & Example
<b>Component ID</b>	INF- BKC -124
<b>Component Name</b>	Blockchain Authenticator
<b>Description</b>	Grants access to specific channel(s) of the blockchain network.
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
<b>INFINITECH Component Category</b>	Blockchain
<b>BDVA Layer</b>	Infrastructure/Data
<b>Input (Required by the Component)</b>	<p>A JSON file</p> <ul style="list-style-type: none"> <li>• Including necessary Hyperledger Fabric channel authentication keys and certificates,</li> <li>• organization identifier and its CA certificates and</li> <li>• participant metadata and other relevant metadata.</li> </ul>
<b>Output (Produced by the Component)</b>	<p>A JSON file</p> <ul style="list-style-type: none"> <li>• Including a unique token confirming that authentication is permitted by the subject entity at the subject time.</li> </ul>
<b>Technology or Platform to be used</b>	Hyperledger
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	Apache 2.0
<b>Other Information / Remarks</b>	<ul style="list-style-type: none"> <li>• Project's official website: <a href="https://www.hyperledger.org/">https://www.hyperledger.org/</a></li> <li>• License: <a href="https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)">https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)</a></li> </ul>
<b>Detailed Documentation</b>	<ul style="list-style-type: none"> <li>• <a href="https://hyperledger.github.io/">https://hyperledger.github.io/</a></li> <li>• <a href="https://github.com/hyperledger/fabric">https://github.com/hyperledger/fabric</a></li> </ul>


	<ul style="list-style-type: none"> <li>• <a href="https://wiki.hyperledger.org/display/fabric">https://wiki.hyperledger.org/display/fabric</a></li> <li>• <a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html</a></li> <li>• <a href="https://www.hyperledger.org/projects/fabric">https://www.hyperledger.org/projects/fabric</a></li> <li>• <a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html</a></li> <li>• <a href="https://wiki.hyperledger.org/display/fabric">https://wiki.hyperledger.org/display/fabric</a></li> <li>• <a href="https://github.com/hyperledger/fabric-chaincode-node">https://github.com/hyperledger/fabric-chaincode-node</a></li> <li>• <a href="https://github.com/hyperledger/blockchain-explorer">https://github.com/hyperledger/blockchain-explorer</a></li> <li>• <a href="https://wiki.hyperledger.org/display/explorer">https://wiki.hyperledger.org/display/explorer</a></li> <li>• Documentation of Input and Output Parameters: <a href="https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html">https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html</a></li> </ul>
--	---

### 3.3.6 Blockchain Encryptor

Attribute	Documentation & Example
<b>Component ID</b>	INF- BKC -125
<b>Component Name</b>	Blockchain Encryptor
<b>Description</b>	Encrypts the clients’ sensitive data within the smart contract using AES256.
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
<b>INFINITECH Component Category</b>	Blockchain
<b>BDVA Layer</b>	Infrastructure/Data
<b>Input (Required by the Component)</b>	Type “string” variables: <ul style="list-style-type: none"> <li>• A 32-byte key is required to encrypt the sensitive data.</li> <li>• The actual plaintext data that is going to be encrypted.</li> </ul>
<b>Output (Produced by the Component)</b>	None
<b>Technology or Platform to be used</b>	Hyperledger
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	Apache 2.0
<b>Other Information / Remarks</b>	<ul style="list-style-type: none"> <li>• Project’s official website: <a href="https://www.hyperledger.org/">https://www.hyperledger.org/</a></li> <li>• License: <a href="https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)">https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)</a></li> </ul>

<b>Detailed Documentation</b>	<ul style="list-style-type: none"> <li>• <a href="https://www.solarwindmsp.com/blog/aes-256-encryption-algorithm">https://www.solarwindmsp.com/blog/aes-256-encryption-algorithm</a></li> <li>• <a href="https://www.atpinc.com/blog/what-is-aes-256-encryption">https://www.atpinc.com/blog/what-is-aes-256-encryption</a></li> <li>• <a href="https://searchsecurity.techtarget.com/definition/Advanced-Encryption-Standard">https://searchsecurity.techtarget.com/definition/Advanced-Encryption-Standard</a></li> </ul>
-------------------------------	---

### 3.3.7 Blockchain Decryptor

Attribute	Documentation & Example
<b>Component ID</b>	INF- BKC -126
<b>Component Name</b>	Blockchain Decryptor
<b>Description</b>	Decrypts data retrieved from the blockchain ledger.
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
<b>INFINITECH Component Category</b>	Blockchain
<b>BDVA Layer</b>	Infrastructure/Data
<b>Input (Required by the Component)</b>	Type “string” variables: <ul style="list-style-type: none"> <li>• Each user’s specific 32-byte key, the same key that is used during encryption.</li> <li>• The ciphertext or encrypted user’s data.</li> </ul>
<b>Output (Produced by the Component)</b>	A JSON struct: <ul style="list-style-type: none"> <li>• The plaintext message which consists of the user’s actual data after the response from the AES256 decryption process.</li> </ul>
<b>Technology or Platform to be used</b>	Hyperledger
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	Apache 2.0
<b>Other Information / Remarks</b>	<ul style="list-style-type: none"> <li>• Project’s official website: <a href="https://www.hyperledger.org/">https://www.hyperledger.org/</a></li> <li>• License: <a href="https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)">https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)</a></li> </ul>
<b>Detailed Documentation</b>	

### 3.3.8 Blockchain Transaction Dataset Preparation Component

Attribute	Documentation & Example
<b>Component ID</b>	INF- BKC -127
<b>Component Name</b>	Blockchain Transaction Dataset Preparation Component
<b>Description</b>	This component will be responsible for retrieving raw transaction blocks from the Bitcoin and Ethereum blockchains and parsing the blocks in order to extract Bitcoin, Ethereum and major token transactions. After retrieving all the blocks up until now, this component will be run periodically (e.g. once a week or as needed) to retrieve newly generated blockchain blocks during the period.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	BOUN
<b>INFINITECH Component Category</b>	
<b>IRA - BDVA Layer</b>	Data Processing Architectures
<b>Input (Required by the Component)</b>	Bitcoin and Ethereum blockchain raw blocks, Major Ethereum ERC20 Token contract addresses (such as USD, EURO stable coin contract addresses).
<b>Output (Produced by the Component)</b>	Scalable Transaction Graph Analysis
<b>Technology or Platform to be used</b>	Blockchain blocks must be retrieved from either nodes attached to the blockchain (e.g. running Parity) or from blockchain data supplier gateways (Google, Infura, Cloudflare). Web3.js Ethereum Javascript API and Consensus abi-decoder for smart contract call data parsing
<b>Part of INFINITECH Core</b>	Will be part of the INFINITECH platform.
<b>License</b>	Open source, MIT license
<b>Other Information / Remarks</b>	N/A
<b>Detailed Documentation</b>	N/A

### 3.3.9 Scalable Transaction Graph Analysis Component

Attribute	Documentation & Example
<b>Component ID</b>	INF- BKC -128
<b>Component Name</b>	Scalable Transaction Graph Analysis Component
<b>Description</b>	This component will be responsible for taking massive bitcoin and ethereum public transaction data. Since transaction graph size massive and growing, it will use parallel algorithms to achieve scalability. It will utilize graph and machine learning algorithms to analyse fraudulent transactions
<b>Icon</b>	N/A

<b>IP Owner &amp; Partner in Charge</b>	BOUN
<b>INFINITECH Component Category</b>	
<b>IRA - BDVA Layer</b>	Data Processing Architectures
<b>Input (Required by the Component)</b>	Bitcoin and Ethereum blockchain raw blocks, Major Ethereum ERC20 Token contract addresses (such as USD, EURO stable coin contract addresses).
<b>Output (Produced by the Component)</b>	Transaction Graph Analysis Results that will be presented to the User Interface for Blockchain Transaction Reports and Visualization Component
<b>Technology or Platform to be used</b>	Parallel processing (high performance computing) technologies. Message Passing Interface (MPI). Graph Partitioning Software such as Scotch and/or Metis. Distributed graph algorithms. Machine Learning
<b>Part of INFINITECH Core</b>	Will be part of the INFINITECH platform
<b>License</b>	Open source MIT License
<b>Other Information / Remarks</b>	N/A
<b>Detailed Documentation</b>	N/A

### 3.3.10 User Interface for Blockchain Transaction Reports and Visualization Component

Attribute	Documentation & Example
<b>Component ID</b>	INF- BKC -129
<b>Component Name</b>	User Interface for Blockchain Transaction Reports and Visualization Component
<b>Description</b>	This component will be responsible for providing user interaction with the Scalable Transaction Graph Analysis component within the bank and collect/manage user as well as annotated blacklisted blockchain addresses . It will utilize OpenAPIs (REST APIs) to submit queries consisting of customer blockchain addresses and blacklists to transaction graph analysis component and generate web based reports and visualization based received results
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	AKTIF
<b>INFINITECH Component Category</b>	
<b>IRA - BDVA Layer</b>	Data Visualization and User Interaction
<b>Input (Required by the Component)</b>	Scalable Transaction Graph Analysis component results
<b>Output (Produced by the Component)</b>	Web interfaces and reports showing customer blockchain addresses to fraudulent transactions if any and web based visualization of transaction graph paths and subgraphs

## D2.5 – Specifications of INFINITECH Technologies - I

<b>Technology or Platform to be used</b>	OpenAPIs / REST APIs, Web servers, Javascript, Vis.js graph drawing library (community version) <a href="https://github.com/visjs-community/visjs-network">https://github.com/visjs-community/visjs-network</a>
<b>Part of INFINITECH Core</b>	Will be part of the INFINITECH platform
<b>License</b>	Open source MIT License
<b>Other Information / Remarks</b>	N/A
<b>Detailed Documentation</b>	N/A

### 3.4 Component Group: Data Semantics

#### 3.4.1 Semantic Stream Analyzer

TBD in future deliverable release D2.6

#### 3.4.2 Semantic Reasoner

TBD in future deliverable release D2.6

#### 3.4.3 Ontology Mapping

TBD in future deliverable release D2.6

#### 3.4.4 Preprocessing

TBD in future deliverable release D2.6

#### 3.4.5 Anomaly Analysis

Attribute	Documentation & Example
<b>Component ID</b>	INF-DSM-130
<b>Component Name</b>	Anomaly Analysis
<b>Description</b>	Anomaly Analysis provides two main functionalities: <ul style="list-style-type: none"> <li>· Anomaly detection</li> <li>· Anomaly prediction</li> </ul> for time series data.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	JSI
<b>INFINITECH Component Category</b>	
<b>IRA - BDVA Layer</b>	Data Analytics
<b>Input (Required by the Component)</b>	Data Preprocessing
<b>Output (Produced by the Component)</b>	User interaction
<b>Technology or Platform to be used</b>	QMiner
<b>Part of INFINITECH Core</b>	Will be part of the INFINITECH platform
<b>License</b>	Open source MIT License
<b>Other Information / Remarks</b>	N/A
<b>Detailed Documentation</b>	N/A

## 3.4.6 Pattern Analysis

Attribute	Documentation & Example
<b>Component ID</b>	INF- DSM -131
<b>Component Name</b>	Pattern Analysis
<b>Description</b>	<p>Pattern Analysis provides two main functionalities:</p> <ul style="list-style-type: none"> <li>· Pattern matching</li> <li>· Discovery</li> </ul> <p>The component will provide support for detection of complex patterns on data graphs</p>
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	JSI
<b>INFINITECH Component Category</b>	Analytics
<b>IRA - BDVA Layer</b>	Data Analytics
<b>Input (Required by the Component)</b>	Data Preprocessing
<b>Output (Produced by the Component)</b>	User interaction
<b>Technology or Platform to be used</b>	SNAP, QMiner
<b>Part of INFINITECH Core</b>	Yes
<b>License</b>	Open source, MIT License
<b>Other Information / Remarks</b>	N/A
<b>Detailed Documentation</b>	N/A

## 3.4.7 Stream Story

Attribute	Documentation & Example
<b>Component ID</b>	INF- DSM -132
<b>Component Name</b>	Stream Story
<b>Description</b>	<p>Stream Story is a component for the analysis of multivariate time series. It computes and visualizes a hierarchical Markov chain model which captures the qualitative behaviour of the systems' dynamics, where system is described with a group of timeseries</p>
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	JSI



<b>INFINITECH Component Category</b>	
<b>IRA - BDVA Layer</b>	Data Analytics
<b>Input (Required by the Component)</b>	Data Preprocessing
<b>Output (Produced by the Component)</b>	User interaction
<b>Technology or Platform to be used</b>	Stream Story, QMiner
<b>Part of INFINITECH Core</b>	Will be part of the INFINITECH platform
<b>License</b>	Open source, MIT License
<b>Other Information / Remarks</b>	N/A
<b>Detailed Documentation</b>	N/A

### 3.5 Component Group: Internet of Things

#### 3.5.1 IoT Registry

Components extracted from the ATOS' Connected Car Framework

Attribute	Documentation & Example
<b>Component ID</b>	INF-IOT-133
<b>Component Name</b>	IoT Context Management and Historical data component
<b>Description</b>	A FIWARE-Based framework designed to capture, process and distribute IoT data. Customised and specialized in real time road traffic and vehicle's information. It will implement Pub/Sub mechanisms and Geolocation and Time series tools.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ATOS
<b>INFINITECH Component Category</b>	
<b>IRA - BDVA Layer</b>	Data Processing Architecture
<b>Input (Required by the Component)</b>	Inputs from Data Management and Data protection layers (Nodes providing Traffic and Vehicles datasets): <ul style="list-style-type: none"> <li>● Standard/DataModel: based on NGSI-v2 FIWARE data models</li> <li>● Format: JSON</li> <li>● Procotols: HTTP/HTTPS/MQTT</li> </ul>
<b>Output (Produced by the Component)</b>	Provides information to Data Analytics components, as the AI framework (EASIER.AI), Real Time analytics nodes and End-user dashboards.

	<ul style="list-style-type: none"> <li>● Standard/DataModel: based on NGSI-v2 FIWARE data models</li> <li>● Format: JSON</li> <li>● Procotols: HTTP/HTTPS/MQTT</li> </ul>
<b>Technology or Platform to be used</b>	<p>Customised technological building blocks based on FIWARE architecture:</p> <ul style="list-style-type: none"> <li>● Context Broker (Orion) [NGSI &amp; ETSI NGSI-LD]: distribute context data between any possible combination of data-producers and data-consumers</li> <li>● Historical Data (Quantumleap) [ETSI NGSI-LD]: persistence of the data managed by the Context Broker.</li> </ul>
<b>Part of INFINITECH Core</b>	Offered as part of the INFINITECH platform
<b>License</b>	Context Broker – Orion: AGPL-v3 Historical Data - Quantumleap: MIT License
<b>Other Information / Remarks</b>	N/A
<b>Detailed Documentation</b>	N/A

### 3.5.2 IoT Stream Handler

Attribute	Documentation & Example
<b>Component ID</b>	INF- IOT -134
<b>Component Name</b>	IoT Agent
<b>Description</b>	A FIWARE-Based component designed to capture device data, homogenise its information and inject the data streaming into the IoT Context Management framework using their own native protocols. IoT Agents should also be able to deal with security aspects of the FIWARE platform (authentication and authorization of the channel) and provide other common services to the device programmer.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ATOS
<b>INFINITECH Component Category</b>	
<b>IRA - BDVA Layer</b>	Data Processing Architecture
<b>Input (Required by the Component)</b>	<p>Current IoT Agents implementations include the following:</p> <ul style="list-style-type: none"> <li>● IoTAgent-JSON - a bridge between HTTP/MQTT messaging (with a JSON payload) and NGSI</li> <li>● IoTAgent-LWM2M - a bridge between the Lightweight M2M protocol and NGSI</li> <li>● IoTAgent-UL - a bridge between HTTP/MQTT messaging (with an UltraLight2.0 payload) and NGSI</li> </ul>

	<ul style="list-style-type: none"> <li>● IoTagent-LoRaWAN - a bridge between the LoRaWAN protocol and NGSI</li> </ul>
<b>Output (Produced by the Component)</b>	<p>All IoT Agents covered here provide their outcomes in the same way:</p> <ul style="list-style-type: none"> <li>● Standard/DataModel: based on NGSI-v2 FIWARE data models</li> <li>● Format: JSON</li> <li>● Procotols: HTTP/HTTPS/MQTT</li> </ul>
<b>Technology or Platform to be used</b>	<p>Customised technological building blocks based on FIWARE architecture:</p> <ul style="list-style-type: none"> <li>● IoT-Agents [NGSI, MQTT, HTTP, UL2.0, JSON]: act as an intermediate between a data-producer (IoT, data-sources, streams) and the Context Broker. Translate data and protocols to the common data model and protocol of the Context Broker. Different IoT-Agents will be used depending on the protocols and data to manage as inputs.</li> </ul>
<b>Part of INFINITECH Core</b>	Offered as part of the INFINITECH platform
<b>License</b>	IoT-Agents: It will depend on the agent, but these are encouraged to use AGPL. In any case, OpenSource IoT Agent
<b>Other Information / Remarks</b>	N/A
<b>Detailed Documentation</b>	N/A

## 3.6 Component Group: Machine Learning

### 3.6.1 Fraud Detection Service Training

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -135
<b>Component Name</b>	Fraud Detection Service Training
<b>Description</b>	This component trains the Machine Learning model that will classify the drivers' routes according to the way they drive (speed, acceleration, braking...)
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ATOS
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	A historic of drivers' data captured on each route they drive.

<b>Output</b>	A trained model that will classify the drivers' way of driving on each route.
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	TensorFlow, Keras, Docker, SKLearn
<b>License</b>	ATOS Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 11 "Personalized insurance products based on IoT connected vehicles").
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.2 Fraud Detection Service Execution

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -136
<b>Component Name</b>	Fraud Detection Service Execution
<b>Description</b>	This component classifies the drivers' routes according to the way they drive (speed, acceleration, braking...) using the model trained in component 3.6.1.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ATOS
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	Drivers' data captured on one route made by the car.
<b>Output</b>	A profile of the driver for the particular route provided.
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	TensorFlow, Keras, Docker, SKLearn
<b>License</b>	ATOS Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 11 "Personalized insurance products based on IoT connected vehicles").
<b>Part of Regulatory Compliance Tool?</b>	No

<b>Other Information / Remarks</b>	TBD
------------------------------------	-----

### 3.6.3 Pay As You Drive Service Training

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -137
<b>Component Name</b>	Pay As You Drive Service Training
<b>Description</b>	This component trains the Machine Learning model that will classify the drivers' behaviour while driving according to the data collected from their car
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ATOS
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	A historic of drivers' data captured from their car, labeled as good or bad drivers.
<b>Output</b>	A trained model that will classify the driver's behaviour as good or bad.
<b>Already available</b>	Yes
<b>Enhancements needed</b>	To be enhanced with INFINITECH Pilot 11 data sources
<b>Technology to be used</b>	TensorFlow, Keras, Docker, SKLearn
<b>License</b>	ATOS Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 11 "Personalized insurance products based on IoT connected vehicles").
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.4 Pay As You Drive Service Execution

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -138
<b>Component Name</b>	Pay As You Drive Service Execution
<b>Description</b>	This component will classify the drivers' behaviour while driving according to the data collected from their car.

<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ATOS
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	Driver's data captured from their car from different trips.
<b>Output</b>	Driver's classified behaviour while driving.
<b>Already available</b>	Yes
<b>Enhancements needed</b>	To be enhanced with INFINITECH Pilot 11 data sources
<b>Technology to be used</b>	TensorFlow, Keras, Docker, SKLearn
<b>License</b>	ATOS Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 11 "Personalized insurance products based on IoT connected vehicles").
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.5 Recommendation Engine Training

TBD in future deliverable release D2.6

### 3.6.6 Recommendation Engine Execution

TBD in future deliverable release D2.6

### 3.6.7 Association Rules Training

TBD in future deliverable release D2.6

### 3.6.8 Association Rules Execution

TBD in future deliverable release D2.6

### 3.6.9 Unexpected transaction detector

Attribute	Documentation & Example
<b>Component ID</b>	INF-MLE-139
<b>Component Name</b>	Unexpected transaction detector
<b>Description</b>	This component identifies unexpected transaction behaviour of the customer based on their prior transactions and forwards them in the form of alerts to the Smart Alert component at real-time.

<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LIB, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	Account transaction updates component
<b>Output</b>	An alert published in the form of JSON/XML file containing the identified alerts through the Smart Alert topic
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	PySpark, Apache Flink, Keras, SKLearn
<b>License</b>	LIB & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5a “Smart and Personalized Pocket Assistant for Personal Financial Management”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.10 Account transaction forecaster

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -140
<b>Component Name</b>	Account transaction forecaster
<b>Description</b>	This component generates predictions about the transaction the customer needs to do based on their transaction behavioural patterns
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LIB, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	Customer’s Account transactions history Customer’s Account transactions updates
<b>Output</b>	A JSON/CSV/XML file containing the forecasted transactions that will be stored in the Account transaction forecaster DB
<b>Already available</b>	No

<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	SKLearn, PySpark, Keras
<b>License</b>	LIB & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5a “Smart and Personalized Pocket Assistant for Personal Financial Management”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.11 Overdraft forecaster

<b>Attribute</b>	<b>Documentation &amp; Example</b>
<b>Component ID</b>	INF- MLE -141
<b>Component Name</b>	Overdraft forecaster
<b>Description</b>	This component is used to predict the overdraft of the customer based on the outcomes of the P5a_Analytics_02 component which are stored in the Account transaction forecaster store
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LIB, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	P5a_Analytics_02 Account transaction forecaster DB
<b>Output</b>	Publish to the Smart Alerts topic insights in the form of JSON/XML regarding the customer’s overdraft
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	SKLearn, PySpark, Keras, Apache Flink
<b>License</b>	LIB & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5a “Smart and Personalized Pocket Assistant for Personal Financial Management”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD



## 3.6.12 Automation proposer

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -142
<b>Component Name</b>	Automation proposer
<b>Description</b>	This component is based on the forecast of the P5a_Analytics_02 component and its role is to provide the customer recommendations upon scheduled manual transactions
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LIB, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	P5a_Analytics_02 and Account transaction forecaster DB
<b>Output</b>	Publish to the Smart Automation Alerts topic insights in the form of JSON/XML regarding the customer’s manual repeated transactions
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	SKLearn, PySpark, Keras
<b>License</b>	LIB & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5a “Smart and Personalized Pocket Assistant for Personal Financial Management”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

## 3.6.13 Transaction sentinel

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -143
<b>Component Name</b>	Transaction sentinel
<b>Description</b>	This component collects information related to the customer’s account transactions and provides warnings and alerts in cases where additional fees or commissions are included in a transaction, or invoices are billed higher than usual.
<b>Icon</b>	<b>N/A</b>

<b>IP Owner &amp; Partner in Charge</b>	LIB, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	Account transaction updates component
<b>Output</b>	Alerts and warnings related to additional fees or commissions, or invoices higher than usual. These alerts are forwarded to the customer through the smart sentinel topic
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	SKLearn, PySpark, Keras
<b>License</b>	LIB & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5a “Smart and Personalized Pocket Assistant for Personal Financial Management”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.13 Bank’s product recommender

<b>Attribute</b>	<b>Documentation &amp; Example</b>
<b>Component ID</b>	INF- MLE -144
<b>Component Name</b>	Bank’s product recommender
<b>Description</b>	Generates recommendations to the customer with respect to the bank’s product offerings. These custom recommendations are generated taking under consideration also the transactions of the customer
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LIB, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.

<b>BDVA Layer</b>	Analytics
<b>Input</b>	Customer's Account transactions DB
<b>Output</b>	Recommendation of bank products to be suggested to the customer through the Smart Recommendation topic
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	SKLearn, PySpark, Keras
<b>License</b>	LIB & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5a "Smart and Personalized Pocket Assistant for Personal Financial Management").
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.14 Smart expense advisor

<b>Attribute</b>	<b>Documentation &amp; Example</b>
<b>Component ID</b>	INF- MLE -145
<b>Component Name</b>	Smart expense advisor
<b>Description</b>	This component takes under consideration the customer's profile and provides custom recommendation regarding the upcoming expenses the customer should cover.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LIB, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	Customer profile and Customer's account transactions DB
<b>Output</b>	Recommendations in the form related to the upcoming customer expenses published in the Smart Expense Advice topic
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	SKLearn, PySpark, Keras
<b>License</b>	LIB & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5a "Smart and Personalized Pocket Assistant for Personal Financial Management").


<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.15 Savings proposer

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -146
<b>Component Name</b>	Savings proposer
<b>Description</b>	This component takes under consideration the customer's transaction history and the transaction forecast and provides custom recommendations related to the way the customer should manage their financial resources. In particular, such recommendations could include a contribution to a saving plan when extra liquidity occurs or the change of a service provider when the expenses exceed the income.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	LIB, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	Customer's account transactions DB
<b>Output</b>	Recommendations related to the saving actions that could be performed, published in the Smart Expense Advice topic
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	SKLearn, PySpark, Keras
<b>License</b>	LIB & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5a "Smart and Personalized Pocket Assistant for Personal Financial Management").
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD


### 3.6.16 Recommender

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -147

<b>Component Name</b>	Recommender
<b>Description</b>	Generates an actionable insight depending on the output from the remaining components within the Data Analytics layer, i. e. components P5b_Analytics_02 to P5b_Analytics_08.
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	BOC, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	P5b_Data_04 (Collected Enriched Data) / P5b_Data_02 (Generated Insights)  A JSON/CSV / XML file from components P5b_Data_04 (Collected Enriched Data) / P5b_Data_02 (Generated Insights) directly or indirectly via component P5b_Data_01 (Data Streaming).
<b>Output</b>	P5b_Data_03 (Generated Actions).  A JSON/CSV/XML file containing the obtained insights and recommendation to be provided to the respective SME via component P5b_Data_03 (Generated Actions).
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
<b>License</b>	BOC & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD


### 3.6.17 Cash Flow Prediction Model

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -148
<b>Component Name</b>	Cash Flow Prediction

<b>Description</b>	<p>ML/AI model used to indicate and predict the available working capital (operating cash flow) of the SME (AS-IS &amp; near term future).</p> <p>Alerts/notifications (via BOC Middleware PushNotifications) to be pushed to the respective SME in case of potential lack of liquidity and/or balance moving below a threshold.</p> <p>Cashflow Data/Insights to be provided to BOC Middleware Mobile/Web BPF.</p> <p>In order to provide valuable insights to the SME the data should be collected and streamed in real (or near real) time, since whenever a new object/entry appears the model should retrain and adapt.</p>
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	BOC, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics.
<b>Input</b>	<p>P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data).</p> <p>A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming).</p> <p>Data required will originate from the following data sources:</p> <ul style="list-style-type: none"> <li>· Transaction (Txn.) Data from the Bank.</li> <li>· Transaction (Txn.) Data from Open Banking.</li> <li>· Transaction (Txn.) Data from SME.</li> </ul>
<b>Output</b>	<p>P5b_Data_02 (Generated Insights).</p> <p>Operating Cash Flow statement (flows, values, time) matrix represent history, actual and near-term future prediction.</p> <p>Matrix will be updated in real (or near real) time to provide accurate insights.</p> <p>Matrix will allow parameterization based on time period and interval (e.g. weeks, months).</p> <p>A JSON/CSV / XML file containing the obtained cash flow related data to be provided to SME directly from component P5b_Data_04</p>

	(Collected Enriched Data) or indirectly via component P5b_Data_01 (Data Streaming).
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
<b>License</b>	BOC & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.18 Budget Prediction


Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -149
<b>Component Name</b>	Budget Prediction
<b>Description</b>	<p>AI (ML) model is used to support the budget target setting for the various categories used by the respective SME. Doing so by providing budget predictions for each utilized category.</p> <p>The underlying model will take into consideration the cash flow analysis output, benchmark, macroeconomic and other available SME data (Business Plan).</p>
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	BOC, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	<p>P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data) / P5b_Data_05 (SME Input).</p> <p>A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming).</p> <p>Data required will originate from the following data sources:</p>

	<ul style="list-style-type: none"> <li>● Transaction (Txn.) Data from the Bank.</li> <li>● Transaction (Txn.) Data from Open Banking.</li> <li>● Transaction (Txn.) Data from SME.</li> <li>● Customer (SME) Data from the Bank.</li> <li>● Other Data from SME.</li> <li>● SME Input from Bank (Mobile app).</li> <li>● Other Data from market (e.g. Benchmarking &amp; Macroeconomic).</li> </ul> <p>and engines</p> <ul style="list-style-type: none"> <li>● Cash Flow Prediction</li> </ul> <p>Benchmark(s)</p>
<b>Output</b>	<p>P5b_Data_02 (Generated Insights).</p> <p>A JSON/CSV / XML file containing the derived budget target for each category used by the respective SME. Output to be provided to SME via component P5b_Data_02 (Insights) / P5b_Data_04 (Collected Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).</p>
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
<b>License</b>	BOC & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”)
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.19 KPI Engine


Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -150
<b>Component Name</b>	KPI Engine
<b>Description</b>	The KPI engine calculates KPIs with regards to the Financial Health and Performance of the respective SME.



	Doing so by taking into consideration the respective SME profile (e.g. maturity stage), accounting-wise optimal KPI values and how other similar SMEs perform.
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	BOC, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms
<b>BDVA Layer</b>	Analytics
<b>Input</b>	<p>P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data) / P5b_Data_05 (SME Input).</p> <p>A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming).</p> <p>Data required will originate from the following data sources:</p> <ul style="list-style-type: none"> <li>● Account (Acc.) Data from the Bank.</li> <li>● Account (Acc.) Data from Open Banking.</li> <li>● Customer (SME) Data from the Bank.</li> <li>● Other Data from SME.</li> <li>● SME Input from Bank (Mobile app).</li> <li>● Other Data from market (e.g. Benchmarking &amp; Macroeconomic).</li> <li>● and engines</li> <li>● Invoices Processing</li> </ul> <p>Benchmark(s)</p>
<b>Output</b>	<p>P5b_Data_02 (Generated Insights).</p> <p>A JSON/CSV / XML file containing results on Financial Health and Performance matrix. Output to be provided to SME via component P5b_Data_02 (Insights) / P5b_Data_04 (Collected Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).</p>
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
<b>License</b>	BOC & UPRC Proprietary.


<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.20 Transaction (Txn) Monitoring

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -151
<b>Component Name</b>	Transaction (Txn) Monitoring
<b>Description</b>	<p>A dynamic complex event processing (CEP) mechanism that monitors the transactions of the user. In case transaction amount or type deviates from normal behaviour the user will be informed of abnormal transactions in order to be safeguarded from double payment mistakes and potential fraud attempts.</p> <p>In addition, expense pattern are also analysed to identify potential savings for instance multiple subscription spending or high ATM fees.</p>
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	BOC, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms
<b>BDVA Layer</b>	Analytics
<b>Input</b>	<p>P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data) / P5b_Data_05 (SME Input).</p> <p>A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming).</p>
<b>Output</b>	<p>P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data).</p> <p>A JSON/CSV / XML file containing results on abnormal transactions and suspicious expenses. Output to be provided to SME via component P5b_Data_02 (Insights) / P5b_Data_04 (Collected</p>

	Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
<b>License</b>	BOC & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD


### 3.6.21 Transaction (Txn) Categorization

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -152
<b>Component Name</b>	Transaction (Txn) Categorization
<b>Description</b>	<p>Smart transaction auto-classification which would also allow the user to manually override the given transaction category and define a new one (re-classify).</p> <p>The categorization performed will be based on the needs of the individual SME.</p>
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	BOC, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms
<b>BDVA Layer</b>	Analytics
<b>Input</b>	P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data) / P5b_Data_05 (SME Input).


	<p>A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming).</p> <p>Data required will originate from the following data sources:</p> <ul style="list-style-type: none"> <li>• Transaction (Txn.) Data from the Bank.</li> <li>• Transaction (Txn.) Data from Open Banking.</li> <li>• Customer (SME) Data from the Bank.</li> <li>• Other Data from SME.</li> </ul> <p>§ SME Input from Bank (Mobile app).</p>
<b>Output</b>	<p>P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data).</p> <p>Output to be provided to SME via component P5b_Data_02 (Insights) / P5b_Data_04 (Collected Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).</p>
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
<b>License</b>	BOC & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.22 Invoice Processing

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -153
<b>Component Name</b>	Invoice Processing
<b>Description</b>	<p>Processes transaction data and ERP data in order to keep record of the invoices which have been partially or fully paid by the SME. Provides insights regarding the respective VAT amount payable and the optimization of cash flow by providing background info on invoices such as paying a invoice at the “right” time.</p> <p>In cases where available invoice data is limited the engine will utilize a simplified approach in order to derive the expected VAT amount to be paid by the SME at the next VAT due date. Doing so by utilizing the banks transaction info or/and past VAT payment amounts.</p>

<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	BOC, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	<p>P5b_Data_04 (Collected Enriched Data).</p> <p>A JSON/CSV / XML file directly from component P5b_Data_04 (Collected Enriched Data) or indirectly via component P5b_Data_01 (Data Streaming).</p> <p>Data required will originate from the following data sources:</p> <ul style="list-style-type: none"> <li>§ Transaction (Txn.) Data from the Bank.</li> <li>§ Transaction (Txn.) Data from Open Banking.</li> <li>§ Customer (SME) Data from the Bank.</li> <li>§ Other Data from SME.</li> </ul> <p>§ SME Input from Bank (Mobile app).</p>
<b>Output</b>	<p>P5b_Data_02 (Insights) and P5b_Data_04 (Data Repository).</p> <p>A JSON/CSV / XML file containing Matrix with invoice information and payment prioritization. Output to be provided to other component P5b_Analytics_02 (Cash Flow Prediction) and SME via component P5b_Data_02 (Insights) and P5b_Data_04 (Collected Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).</p>
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	TBD
<b>License</b>	BOC & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).
<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

## 3.6.23 Benchmark(s)

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -154
<b>Component Name</b>	Benchmark(s)
<b>Description</b>	<p>Generates benchmarks to allow the SME to compare to likewise businesses (same industry, maturity stage, etc.).</p> <p>This is done in relation to component P5b_Analytics_03 (Budget Prediction), P5b_Analytics_04 (KPI Engine) and P5b_Analytics_07 (Invoice Processing).</p>
<b>Icon</b>	
<b>IP Owner &amp; Partner in Charge</b>	BOC, UPRC (Business and ML Model).
<b>INFINITECH Component Category</b>	Analytics & ML Algorithms.
<b>BDVA Layer</b>	Analytics
<b>Input</b>	<p>P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data) / P5b_Data_05 (SME Input).</p> <p>A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming).</p> <p>Data required will originate from all data sources.</p>
<b>Output</b>	<p>P5b_Data_02 (Generated Insights).</p> <p>A JSON/CSV / XML file containing Benchmark Matrix. Output to be provided to SME via component P5b_Data_02 (Insights) and P5b_Data_04 (Collected Enriched Data) directly or via component P5b_Data_01 (Data Streaming).</p>
<b>Already available</b>	No
<b>Enhancements needed</b>	To be implemented and deployed.
<b>Technology to be used</b>	Python libraries (Pandas, Scikit-learn, Tensorflow, Keras).
<b>License</b>	BOC & UPRC Proprietary.
<b>Part of INFINITECH Core?</b>	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).

<b>Part of Regulatory Compliance Tool?</b>	No
<b>Other Information / Remarks</b>	TBD

### 3.6.24 KMeans

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -155
<b>Component Name</b>	KMeans
<b>Description</b>	Batch BDA Service: Given a set of observations (x1, x2, ..., xn), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k ( $\leq n$ ) sets $S = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster variance.
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ENG
<b>INFINITECH Component Category</b>	ML Algorithm
<b>IRA - BDVA Layer</b>	Analytics/User
<b>Input (Required by the Component)</b>	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE Input type: Quantitative
<b>Output (Produced by the Component)</b>	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE
<b>Technology or Platform to be used</b>	It is a BDA service registered in the ALIDA catalogue. It is a PySpark-based micro-service running on K8S cluster mode.
<b>Part of INFINITECH Core</b>	This component is general purpose and reusable as a BDA service within ALIDA
<b>License</b>	Component developed with open source technologies. ENG Proprietary
<b>Other Information / Remarks</b>	<p>In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies.</p> <p>This service is registered in ALIDA catalogue as Spring Boot Application containing the python code and its dependencies. After implementing the algorithm using Pyspark, creating the Dockerfile and pushing the new image inside a repository, this microservice is registered into the ALIDA catalogue through the GUI.</p> <p>Source: <a href="https://home.alidalab.it/">https://home.alidalab.it/</a></p>

<b>Detailed Documentation</b>	N/A
-------------------------------	-----

### 3.6.25 Random Forest (Model)

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -156
<b>Component Name</b>	Random Forest (Model)
<b>Description</b>	Batch BDA Service: An ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ENG
<b>INFINITECH Component Category</b>	ML Algorithm
<b>IRA - BDVA Layer</b>	Analytics/User
<b>Input (Required by the Component)</b>	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE
<b>Output (Produced by the Component)</b>	model object stored into HDFS.
<b>Technology or Platform to be used</b>	It is a BDA service registered in the ALIDA catalogue. It is a micro- PySpark-based service running on K8S cluster mode.
<b>Part of INFINITECH Core</b>	This component is general purpose and reusable as a BDA service within ALIDA.
<b>License</b>	Component developed with open source technologies. ENG Proprietary
<b>Other Information / Remarks</b>	<p>In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies.</p> <p>This service is registered in ALIDA catalogue as Spring Boot Application containing the python code and its dependencies. After implementing the algorithm using Pyspark, creating the Dockerfile and pushing the new image inside a repository, this microservice is registered into the ALIDA catalogue through the GUI.</p> <p>Source: <a href="https://home.alidalab.it/">https://home.alidalab.it/</a></p>



<b>Detailed Documentation</b>	N/A
-------------------------------	-----

### 3.6.26 Random Forest (Predict)

Attribute	Documentation & Example
<b>Component ID</b>	INF- MLE -157
<b>Component Name</b>	Random Forest (Predict)
<b>Description</b>	Streaming random forests algorithm
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ENG
<b>INFINITECH Component Category</b>	ML Algorithm
<b>IRA - BDVA Layer</b>	Analytics/User
<b>Input (Required by the Component)</b>	This component is a consumer who subscribes to Kafka's topic, so it reads the messages that will be sent to the topic. Useful information to take stored machine learning model created by Random Forest Model.
<b>Output (Produced by the Component)</b>	This component is also a producer, who write the results to other one Kafka topic.
<b>Technology or Platform to be used</b>	It will be a BDA service registered in the ALIDA catalogue
<b>Part of INFINITECH Core</b>	This component is general purpose and reusable as a BDA service within ALIDA
<b>License</b>	Component developed with open source technologies. ENG Proprietary
<b>Other Information / Remarks</b>	In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies. This service is registered in ALIDA catalogue as Spring Boot Application containing the python code and its dependencies. After implementing the algorithm using Pyspark, creating the Dockerfile and pushing the new image inside a repository, this microservice is registered into the ALIDA catalogue through the GUI. Source: <a href="https://home.alidalab.it/">https://home.alidalab.it/</a>
<b>Detailed Documentation</b>	N/A

### 3.7 Component Group: Visualization

#### 3.7.1 Data Visualization Dashboard

TBD in a future edition of this deliverable D2.6

#### 3.7.2 Model Visualization Dashboard

TBD in a future edition of this deliverable D2.6

#### 3.7.3 Results Visualization Dashboard

TBD in a future edition of this deliverable D2.6

#### 3.7.4 Real-Time Visualization

Attribute	Documentation & Example
<b>Component ID</b>	INF-VIS-158
<b>Component Name</b>	Real Time Visualization
<b>Description</b>	Useful tool for displaying charts through web application
<b>Icon</b>	N/A
<b>IP Owner &amp; Partner in Charge</b>	ENG
<b>INFINITECH Component Category</b>	Visualization
<b>IRA - BDVA Layer</b>	Interaction/Visualization
<b>Input (Required by the Component)</b>	Metadata to handle resulting datasets as data source and query.  This component maintains a channel between browser and storage using a "sent send events".
<b>Output (Produced by the Component)</b>	Chart properly configured
<b>Technology or Platform to be used</b>	Microservices deployed through Kubernetes <ul style="list-style-type: none"> <li>● D3js</li> <li>● Java Component</li> <li>● Kafka</li> <li>● Docker</li> <li>● PRESTO (Data Source Connector)</li> <li>● HIVE (Data Source Connector)</li> <li>● HDFS (Data Source Connector)</li> <li>● KAFKA (Data Source Connector)</li> <li>● REDIS (Data Source Connector)</li> <li>● CASSANDRA (Data Source Connector)</li> <li>● ELASTICSEARCH (Data Source Connector)</li> </ul>
<b>Part of INFINITECH Core</b>	This component is part of ALIDA core
<b>License</b>	Component developed with open source technologies. ENG Proprietary

<b>Other Information / Remarks</b>	In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies. Source: <a href="https://home.alidalab.it/">https://home.alidalab.it/</a>
<b>Detailed Documentation</b>	N/A

## 4. Conclusions and next steps

Scope of task T2.3 is the specification of BigData/IoT technological building blocks to be developed in the scope of the project, in particular in the areas of data and semantic interoperability, cost-effective real-time BigData analytics, elastic cloud storage, integrated (declarative) data querying, AI/ML algorithms.

In particular, the detailed specifications of a large set of innovative components that are illustrated in Section 3, including specifications about the implementation technologies (e.g., BigData/IoT platforms, AI/ML toolkits, HPC infrastructures) will be used to realize and guide the implementation and integration.

Overall, the INFINITECH pilot requirements reflect the State of the Art of the application of BigData, IoT and AI in the Financial Services and contribute to the latest trends.

Within this deliverable, services related to the landscape of technology and data driven innovation of Financial Services are identified based on the broad use cases and reference scenarios provided by the INFINITECH pilots. Based on existing services and processes (e.g., KYC/KYB, fraud detection, customer service, portfolio management, asset management, usage based insurance) it has been outlined in the pilots' scenarios how these could become more autonomous, personalized and context-aware. In addition the pilots preliminary workflows design have been analyzed to provide the required functional services for their fulfilment.

Finally, a mapping of functional components to the BDVA reference model [4] and preliminary IRA design facilitates the next steps within the task T2.7 .

Based on the outcomes of this deliverable, in the next steps of task T2.3 are:

- Analyze state of the art technologies to complete components specifications
- Complete components surveys and map requirements to components' specification
- Find communalities of the features reducing the number of components needed
- Complete component identification, correct mapping and identification

## References

- [1] INFINITECH Deliverable D2.1 – User Stories and Stakeholders' Requirements - I
- [2] INFINITECH Deliverable D2.3 – Reference Scenarios and Use Cases – I
- [3] INFINITECH Deliverable D2.13 – Reference Architecture – I
- [4] BDVA: “BDV SRIA – European Big Data Value Strategic Research and Innovation Agenda”, Version 4.0, October 2017