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

D6.7 - Sandboxes in Incumbent Testbeds - I

Title	D6.7 - Sandboxes in Incumbent Testbeds - I
Revision Number	3.0 [build#54]
Task reference	T6.4
Lead Beneficiary	GFT
Responsible	Maurizio Megliola
Partners	ABILAB AKTIF BANKIA BOC BOI BOS BPFI CP DWF FTS GFT HPE JSI LIB NBG NUIG PI RB
Deliverable Type	Report
Dissemination Level	PU
Due Date	2020-12-31 [M15]
Delivered Date	2021-03-10
Internal Reviewers	ASSEN NOVA
Quality Assurance	INNOV
Acceptance	Accepted by Coordinator
EC Project Officer	Pierre-Paul Sondag
Programme	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	Role[1]	Author(s)[2]	Section(s)
GFT	Lead Beneficiary	Maurizio Megliola	
GFT	Contributor		0. 1. 2. 3. 3.1. 3.2. 3.3. 3.4. 3.5. 3.6. 3.7. 3.8. 3.9.
HPE	Contributor		2.
BANKIA	Contributor		3.1.
PRIVE	Contributor		3.2.
BOC	Contributor		3.3.
NBG	Contributor		3.4.
CXB	Contributor		3.5.
BOS	Contributor		3.6.
AKTIF	Contributor		3.7.
PI	Contributor		3.8.
ABILAB	Contributor		3.9.
ASSEN	Reviewer		all sections
NOVA	Reviewer		all sections
INNOV	Quality Assurance	Dimitris Drakoulis	

1. Lead Beneficiary, Contributor, Internal Reviewer, Quality Assurance
2. Can be left void

Revison History

Version	Date	Partners	Description
0.1	2020-11-30	GFT	TOC Version
0.2	2020-12-08	GFT	Version for Work in Progress
0.3	2021-01-15	GFT	Updated Version with contributions
0.4	2021-02-10	GFT	Version for Work in Progress
1.0	2021-03-05	GFT	Version for Peer Review
1.1	2021-03-09	GFT	Version after peer review
2.0	2021-03-10	GFT	Version for Quality Assurance
3.0	2021-03-10	GFT	Version for Submission

Executive Summary

This document Deliverable D6.7 describes the first outcomes of Task 6.4 (“Incumbent Testbeds Establishment and Customization”). T6.4 is part of WP6, whose objectives are:

1. To analyze the existing testbeds and specify their enhancements and upgrades;
2. To provide tools and techniques for creating tailored sandboxes based on the selection of proper INFINITECH data assets, technological and regulatory building blocks;
3. To provide a mechanism for integrated management of testbeds’ datasets, based on a continuous integration approach;
4. To establish the testbeds for experimentation and validation, including all relevant sandboxes;
5. To ensure continuous technical support for all testbeds, while establishing processes for certification/standardization of digital finance/insurance solutions.

From the point of view of the deployment, Pilots are defined in terms of Testbeds (infrastructure) and Sandboxes (components). In this context, T6.4 is about the physical development of hardware and software infrastructure of the incumbent testbeds and the blueprint for the development of sandboxes of incumbent organizations.

The current deliverable is a first assessment of the project's Pilots provided by the Pilots' end users and presents a schematic (tabular) description and definition of each Testbed and Sandbox components. The different schematics can be considered as a bill of materials of the resources needed to perform the demonstrators. The provided information can actually be used as input for configurators and cost structures to set up the testbeds and therefore it is extremely valuable to organizations from IT to financial and procurement departments.

As an outcome of this assessment, this deliverable confirms that the first KPI of Objective 6 in WP6 has almost been achieved with 8 testbeds instantiated and one more as a result of the Reference Infrastructure (Pilot 15).

However, this deliverable is a preliminary version and some infrmation are yet to be defined (TBD) or not applicable and/or available (N/A). Other two versions namely Deliverable D6.8 and D6.9, will follow in September 2021 and in June 2022 that will complete the assessment with more detailed information to form the "technical recipes" to build the physical pilots' infrastructures of similar pilots and production environments.

Table of Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Services
BOC	Bank of Cyprus
BOS	Bank of Slovenia
CD	CI/CD Continuous Integration/Continuous Development
CI	Continuous Integration
CPU	Central Processing Unit
DL	Deep Learning
DoA	Description of Action (also DoW, description of Work, PART A of Grant Agreement)
EC2	Elastic Cloud Computing (AWS service)
EKS	Elastic Kubernetes Services (AWS Service)
EU	European Union
HPC	High Performance Computing
IP	Internet Protocol
IT	Information Technology
IoT	Internet of Things
KPI	Key Performance Indicator
ML	Machine Language
MPI	Message Passing Interface
MVP	Minimum Viable Product|Platform
N/A	Not Available / Not Applicable
OSS	Open Source Software
POC	Proof of Concept
RA	Reference Architecture
REST	Representational State Transfer
SFTP	Secure File Transfer Protocol
SOAP	Simple Object Access Protocol
SSD	Solid State Drive
TBD	To Be Determined
UI	User Interface
VAT	Value Added Tax
VM	Virtual Machine

1 Introduction

This deliverable belongs to T6.4 (“Incumbent Testbeds Establishment and Customization”), and consequently to WP6 (“Tailored Sandboxes and Testbeds for Experimentation and Validation”). In general, WP6 aims to:

1. Analyze the existing testbeds and specify their enhancements and upgrades;
2. Provide tools and techniques for creating tailored sandboxes based on the selection of proper INFINITECH data assets, technological and regulatory building blocks;
3. Provide a mechanism for integrated management of testbeds’ datasets, based on a continuous integration approach;
4. Establish the testbeds for experimentation and validation, including all relevant sandboxes;
5. Ensure continuous technical support for all testbeds, while establishing processes for certification/standardization of digital finance/insurance solutions.

INFINITECH will provide 15 testbeds for experimentation, testing and validation of BigData and IoT applications in the financial and insurance sectors, including:

- Testbeds managed by incumbent financial organizations of the consortium;
- EU-WIDE testbed that will be made available to Financial/FinTech/InsuranceTech enterprises of the consortium for their pilots;
- Reference testbed available for testing and open pilots.

In this context, the objective of T6.4 is to develop the hardware and software infrastructure of the incumbent testbeds and the blueprint for the development of sandboxes of incumbent organizations. The task will result in a number of sandboxes that will be configured according to the continuous integration / DevOps approach.

Some information are yet to be defined (TBD) or not applicable and/or available (N/A). Other two versions of the current document (D6.8 and D6.9) will follow in September 2021 and in June 2022.

1.1 Objective of the Deliverable

D6.7 describes the results of Task 6.4 for the first part of the project. In particular, it documents the hardware and software infrastructure of the established sandboxes in the infrastructure of incumbent organizations. This document constitutes a report of:

1. A design description and a deployment strategy description of the incumbent infrastructures in working condition: specific deployments, K8s based, according to the INFINITECH general strategy for testbeds and sandboxes, as reported in D6.4 (“Tools and Techniques for Tailored Sandboxes and Management of Datasets - I”);
2. A description of the MVP pilots local testbeds and sandboxes for the first project iteration;
3. An initial mapping of testbeds and sandboxes from 2. to 1.

1.2 Insights from other Tasks and Deliverables

The following figure shows the relationships among the different tasks of WP6.

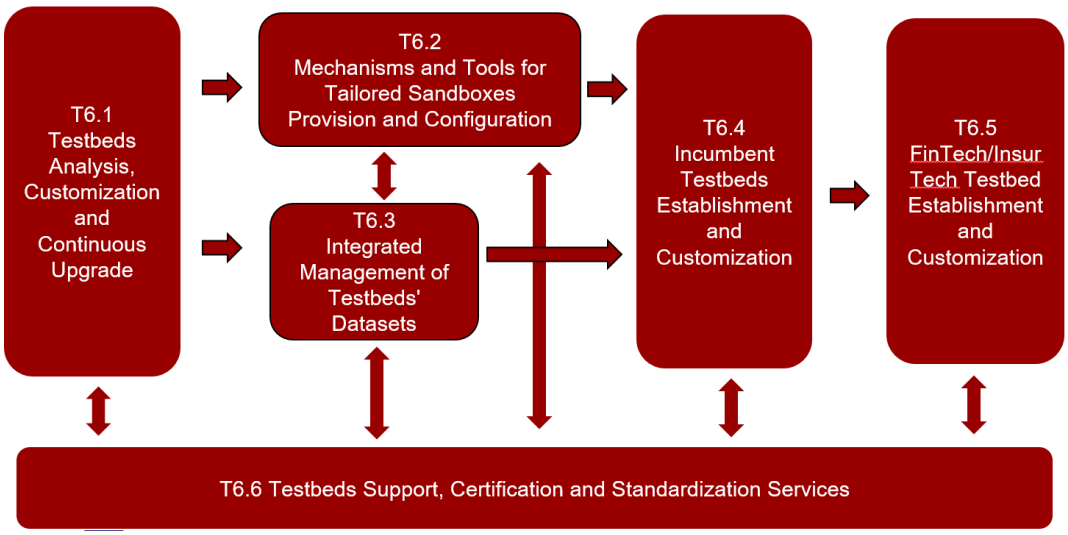

Figure 1 - WP6 internal dependencies per task level

T6.4 leverages mechanisms and tools specified in task 6.2 to create the tailored sandboxes in the testbeds of incumbent financial organizations. As a consequence, D6.4 represents an important input for this document. Moreover, the latter leverages the information on pilots collected in D7.1 (“Report on Pilot Sites Preparation - I”).

1.3 Structure

This deliverable is composed of the following sections:

- Section 1 is the introduction to the deliverable and includes the description of the objective, insights from other tasks and deliverables and the structure;
- Section 2 describes the testbeds and sandboxes implementation general strategy that will be required to be followed from all pilots;
- Section 3 provides a definition of the incumbent and blueprint pilots in terms of testbeds and sandboxes;
- Section 4 concludes the document and provides insights on the future work.

2 Reference Testbeds and Sandboxes

In other project's deliverables, namely D2.x, D6.1, D6.4, D7.1 etc, the basic concepts of developing and deploying the POC of the pilots have been described in details. However, it is important to recall the key concepts underlying the Testbed and Sandbox approach as they serve as a basis to understand the actual stage of the Pilots development. An excerpt from Deliverable D6.4 is reported in this section and the key definitions are as follows.

TESTBED: the ensemble of resources like hardware, storage and networking that support one or more applications. A testbed can be considered from a physical view (the actual configuration of servers, RAM, disks, and network connectivity) or a virtual view (as the set of Virtual Machines and services on top of the physical resources). Usually, this latter view is the one provided by a Cloud Service Provider (like AWS) and also more convenient for a higher level conceptual model of what is needed to support a specific instance of applications.

Therefore, the resources that compose a **TESTBED** can be inside a private Data Centre or in any cloud provider. In INFINITECH, a TESTBED can also be considered logically as the set of virtual resources managed by an orchestrator that deploys on-demand applications for a specific use case. In other words, the infrastructure that provides support for sandboxes.

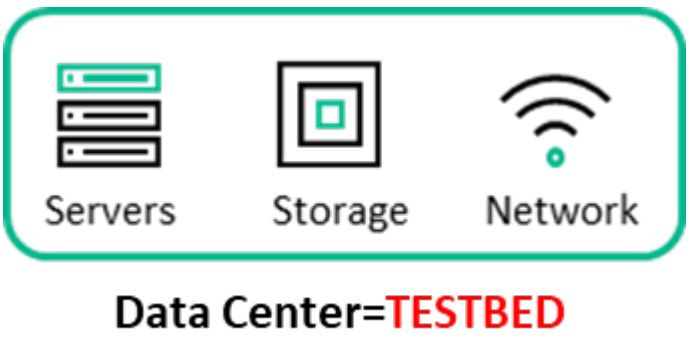

Figure 2 - Testbed

SANDBOX: a group of applications organized to perform a specific task. Sandboxes are deployed and run in the testbed's infrastructure. Sandboxes are a very powerful concept to encapsulate vertical solutions that solves the use cases. Sandboxes are flexible (as the name implies) and can be considered from a high level of abstraction as a configuration of basic components.

Within a Pilot, several use cases can be solved by different sets of applications as in the following figure:

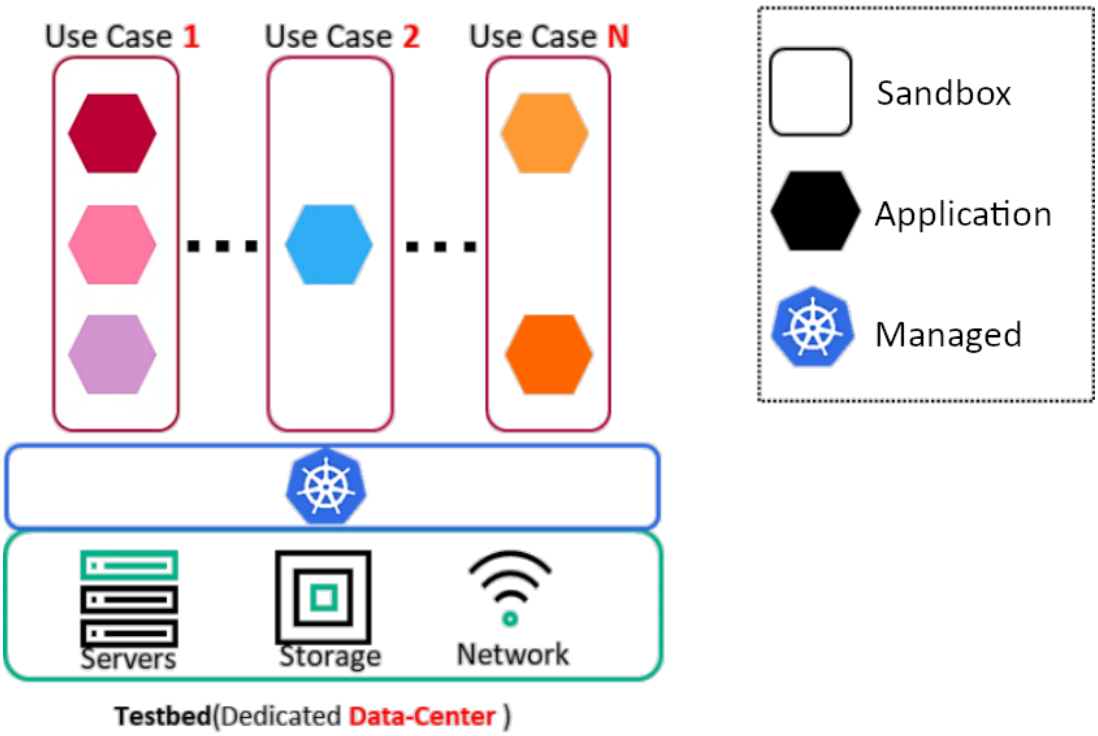

Figure 3 - Testbed vs. Pilot vs. Sandbox

2.1 Reference Blueprint Testbed and Sandboxes

Testbeds and **Sandboxes** have a very specific implementation in the Reference Infrastructure (D6.4) that serves as a blueprint for the whole project. In that respect the logical concepts find their definitions and even instances in the Reference Infrastructure. In the following the technological assumptions are provided.

The **REFERENCE (Blueprint) TESTBED** of the INFINITECH PROJECT is based on a cloud Infrastructure and Services provided by Amazon Web Services or simply AWS.

AWS provides Kubernetes EKS managed services, which we leverage in the INFINITECH blueprint implementation through the provisioning of a simple cluster of two worker nodes:

Testbed	**Components**
Hosting/Cloud	AWS
Server Infrastructure	EC2 Instances
Managed Services	Kubernetes EKS

The Kubernetes Namespace feature makes it possible to logically isolate the different components (called PODs) deployed inside it from the other component deployed in other Namespaces. The Namespace concept is used to provide an INFINITECH sandbox (in the “INFINITECH way”) as it enables network isolation and resources limitation, as depicted in the following picture.

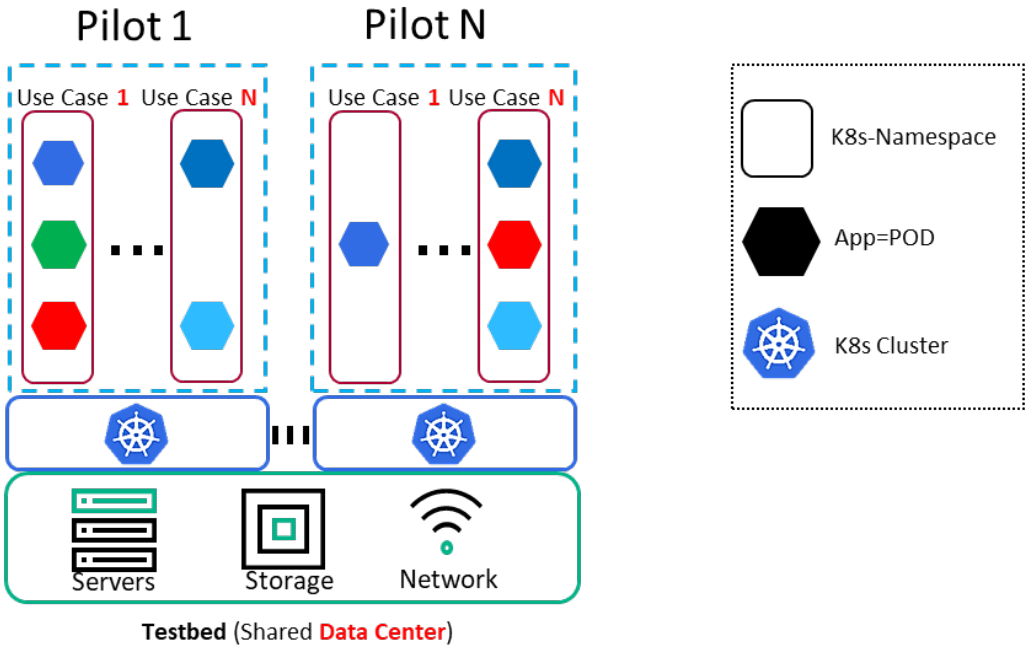

Figure 4 - Sandboxes in a Reference Testbed

2.2 Deployment scenarios and relation with the Reference Testbed and Sandboxes

All incumbent pilots have the opportunity to test some of their specific use cases on the Blueprint Reference Testbed, but the final and target ("production level") pilot will be deployed in a dedicated infrastructure, so at some point it is necessary to plan a migration from the Reference Testbed to the target environment. The following sections provide a high level overview of the possible approaches to migrate in some possible deployment scenarios, which will be detailed in the following versions of this deliverable.

2.2.1 Public Cloud Provider deployment

2.2.1.1 Private AWS account deployment

If the migration target infrastructure is a AWS EKS service (at the time of writing this is the foreseen scenario for Pilots 1, 4, 5b, 9), it is possible to use the Terraform script already used to build the Reference Testbed, in order to replicate the environment. This kind of activity requires very little changes in the configuration files developed for Reference Testbed.

The procedure is well documented on the Terraform website:

https://learn.hashicorp.com/tutorials/terraform/eks

2.2.1.2 Private MS Azure account deployment

If the migration target infrastructure is a MS Azure AKS service (at the time of writing this is the foreseen scenario for Pilot 6), it is possible to use Terraform to replicate the cluster adapting already the developed Terraform script written for AWS to make it working on MS Azure. To understand how to provision an AKS cluster and as consequence how to perform these changes a good start point is the Terraform website:

https://learn.hashicorp.com/tutorials/terraform/aks?in=terraform/kubernetes

2.2.2 On premise deployment

If the migration target infrastructure is located on premise in a incumbent data center (at the time of writing this is the foreseen scenario for Pilots 8 and 10) a specific migration strategy will be required, properly adapting the previous one according to the requirements of the target infrastructure.

2.2.3 Hybrid deployment

The previous sections stated that a testbed can be hosted on private Data Centre or in any cloud provider. However, it is probable that real case scenarios follow a mixed approach: for example, this could be due to legacy reasons or to the fact that a the containerization system used to containerize a specific component is not compatible with the Kubernetes installation in the testbed. Anyway, it is important to find a way to deal with hybrid scenarios. One possible solution is the one already proposed in D6.10, namely the sidecar deployment, which is represented in the following figure.

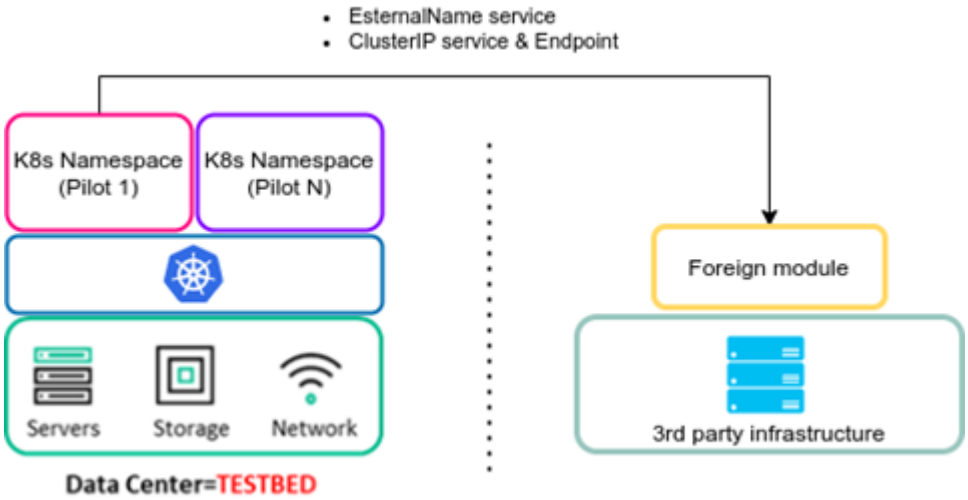

Figure 5 - Sidecar deployment (D6.10)

Following the sidecar deployment, any component can be hosted remotely and managed through to the built-in Kubernetes service-discovering mechanisms and its networking abstraction layer. One approach could be to create a Kubernetes service in the testbed without pod selectors alongside a new Endpoint object that will send the traffic to the target software (see the Kubernetes documentation [1]). Otherwise, if a public hostname is available to access the component in the third party infrastructure, the communication between the sandbox and the external component could be handled through a Kubernates service having type “ExternalName”.

2.2.4 Incumbent Testbeds and Sandboxes

In the following sections the Pilots Testbed and Sandboxes will be described in terms of components. Each Pilot will have its own infrastructure and technology which in principle could be different from all the other and even from the reference implementation. However, the logical concepts stand, and a mapping will be provided by the pilots to confirm the validity of the general approach. In some cases, a planning to adopt the INFINITECH solutions and roadmap is provided as an indication.

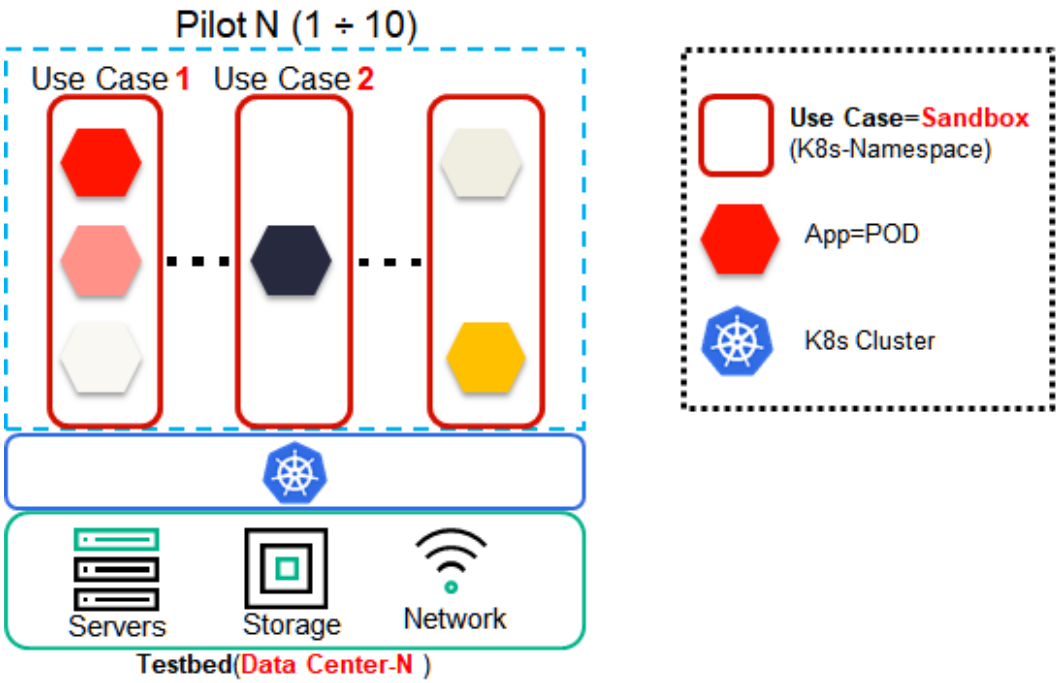

Figure 6 - Roadmap

The schemes suggested for the mapping are in the example tables below

Table 1 - Testbed Mapping

Testbed	Components
Hosting/Cloud	AWS/Azure/On premise
Server Infrastructure	VM with CPU/GPU, RAM, Storage + Operating System e.g. Ubuntu 18.04
Managed Services	Orchestrator e.g. K8s EKS / …
Other tools/Applications	E.g. Kafka, Elastic Search

A typical Pilot will have its own Sandboxes for development (usually following the blueprint DevOps processes) and one or more Sandboxes for the different Use Cases.

Table 2 - Sandbox Table Mapping

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Technological Component Definition	Reference Architecture Mapping Data Source/ Ingestion / Security / Data Management / Analytics / Interface / Presentation) Other tools/ Applications	Developed specifically for Infinitech or legacy or other	Resources: CPU, memory, storage to deploy the component	Container Technology like Docker /…)	Reference Repository (Yes/Not disclosed)	Deployment mode e.g legacy/ manual / Kubernetes	Technology Readiness Level

The different USE CASES are organized in a set of sandboxes. Whenever possible the TRL level of the application used is showed.

A high level figure of the Pilot's Testbed and Sandboxes with a clear mapping to the blueprint Testbed and Sandboxes concepts can clarify the physical and logical view of the Pilot, like in the following example picture.

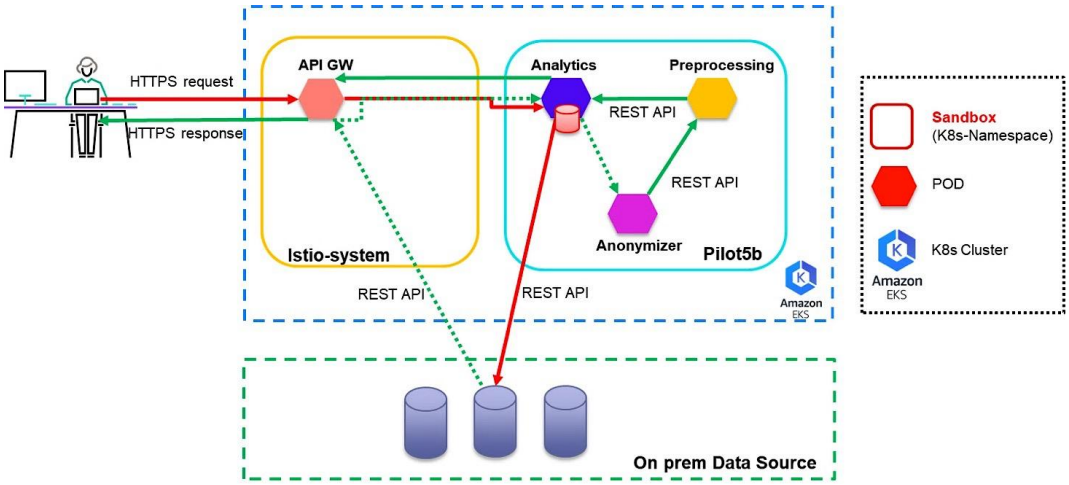

Figure 7 - Pilot physical and logical

A special case of Sandbox is for CI/CD Development. This will be described by the following table.

Table 3 - CI/CD Sandbox Mapping

Sandbox – DevOps (optional)	**Component Details**
Description	
Source code management	
Container	
Orchestrator	
Registry	
CI/CD software	
Other tools/Applications	

3 Testbed and Sandboxes for incumbent pilots

This section defines Pilots 1, 4, 5b, 6, 7, 8, 9, 10, and 15 from a point of view of the deployment, namely in terms of Testbeds (infrastucture) and Sandboxes (components).

3.1 Pilot 1

Pilot 1 ("Invoices Processing Platform for a more Sustainable Banking Industry") is managed by BANKIA. It aims to apply Artificial Intelligence technologies over scanned notary invoices for cost savings and increased effectiveness. AI can be leveraged to extract relevant indicators from digitized invoices: the indicator can then be used to automatically and accurately rate notaries based on a sustainability index. The reader can find more information and a plan for Pilot 4 in Section 2.1 of D7.1.

3.1.1 Testbed Technical Specification

The AWS Bankia Private Cloud will be used to implement a cloud-based testbed. The following table provides further details.

Table 4 - Pilot 1 Testbed

Testbed	Components
Hosting	AWS Private Cloud Bankia
Server Infrastructure	AWS EC2 instance of the type g4dn.xlarge with 200 GB of disk with GPU For inference, normal computing optimized instances c6g.2xlarge or the same type g4dn.xlarge, with the AWS Deep Learning AMI (Ubuntu 18.04).
Managed Services	N/A
Other tools/ Applications	Data management: linux file system, S3, elastic search
	Data processing: kafka, Kubeflow
	Data analytics and AI related tools: tensorflow 1.5, sklearn, pandas, numpy, seaborn
	Data tagging: labelme
	Data visualization: kibana, Floent, Prometheus.

3.1.2 Sandboxes

The following table provides information on the components that will be used within Pilot 1 sandbox.

Table 5 - Pilot 1 Sandbox

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Document pre-processing	Other tools	developed	TBD	Docker	NOT disclosed	Kubernetes	NA
Document entities and region-of-interest extraction	Data Management	developed	TBD	Docker	NOT disclosed	Kubernetes	NA
Entity association	Other tools	developed	TBD	Docker	NOT disclosed	Kubernetes	NA
Business rules engine	Analytics	developed	TBD	Docker	NOT disclosed	Kubernetes	NA
Data Tagger	Data Management	developed	TBD	Docker	NOT disclosed	Kubernetes	NA
Document validator	Analytics	developed	TBD	Docker	NOT disclosed	Kubernetes	NA

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Training and inference orchestrated pipelines.	Other tools	developed	TBD	Docker	NOT disclosed	Kubernetes	NA
MLOps tool	Other tools	developed	TBD	Docker	NOT disclosed	Kubernetes	NA
Reporting business dashboards and operational databases	Presentation	developed	TBD	Docker	NOT disclosed	Kubernetes	NA

3.2 Pilot 4

Pilot 4 ("Personalised Portfolio Management") is managed by PRIVE. It aims to develop and adapt an optimization algorithm and an artificial intelligence engine within the Privé Managers Wealth Management Platform to explore the possibilities of AI Based Portfolio construction for Wealth Management. This will enable the advisor/customer to use the "Prive Managers" Wealth Management Platform and to use its risk-profiling and investment proposal capabilities, starting from her personal risk-awareness. The reader can find more information and a plan for Pilot 4 in Section 2.4 of D7.1.

3.2.1 Testbed Technical Specification

The Testbed for Pilot 4 will be hosted in an infrastructure owned by PRIVE Technologies – Austria. The following table provides further details.

Table 6 - Pilot 4 Testbed

Testbed	Components
Hosting	Privé own Amazon Cloud in AWS
Server Infrastructure	3 VM instances, each with the following: CPU: Intel Xeon 3 GHz or faster Core: minimum 2 Core 4 threads Memory: 32 GB DDR4 1600 or 1866 Hard Disk: 16 GB SSD
Managed Services	N/A
Other tools/ Applications	The SaaS platform runs in multiple data centres with active-active setup to achieve high availability. Privé has the following environments: DEV, SIT, UAT and PROD. Data can be transferred via SFTP, FIX or API. Most Privé APIs are REST, but SOAP and GraphQL are also supported. The architecture is based on microservices, using the following specification: • Operating System: Ubuntu 18.04 LTS • Framework: SpringBoot: 2.2 • Application Server: Tomcat: 7.0.103 • Database: MySQL: 5.6.47 • Database: MongoDB: 3.6 • Language Runtime: Java: OpenJDK 8u242

3.2.2 Sandboxes

The following table provides information on the components that will be used within Pilot 4 sandbox.

Table 7 - Pilot 4 Sandbox

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Data Collection	Data Management	Developed	Not Disclosed	N/A	Not Disclosed	N/A	7

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Customers' & investments/portfolio Data quality check	Data Processing	Developed	Not Disclosed	N/A	Not Disclosed	N/A	7
AI Based Portfolio Optimization Process	Analytics	Developed	Not Disclosed	N/A	Not Disclosed	N/A	N/A

3.3 Pilot 5b

Pilot 5b ("Business Financial Management (BFM) tools delivering a Smart Business Advise") is managed by BOC. It aims to provide Small and Medium sized enterprises (SMEs) clients of Bank of Cyprus with personalized business insights and recommendations on managing the SMEs financial health in the areas of cash flow management, continuous spending/cost analysis, budgeting, revenue review and VAT provisioning. To this aim, the available data will feed a set of AI powered Business Financial Management tools. The reader can find more information and a plan for Pilot 5b in Section 2.5 of D7.1.

3.3.1 Testbed Technical Specification

Pilot 5b testbed will be accommodated by Bank of Cyprus, which is going to provide an AWS environment for the various pilot's components and operation. The following table provides further details.

Table 8 - Pilot 5b Testbed

Testbed	Components	Version
Hosting	Bank of Cyprus (BOC) is developing an AWS testbed, based on the technical requirements and guidelines of the relevant partners, and tailored for the unique pilot's components and the required data ingestion. As the testbed's specifications have not yet been finalised and certain bank processes require time, until the bank's AWS ecosystem is available the pilot's first components will be hosted in GFT's AWS environment.	
Server Infrastructure	20-Core CPU, 64 GB RAM and a GPU with 8 (preferably 16) GB RAM with the ability of GPU-enabled instances for deploying deep learning models.	
Managed Services	NA	
Other tools/ Applications	Baseline Technologies, tools and programming languages	Minimum version required
	Python	3.8
	Docker	18.09.7
	jaydebeapi	1.2.3
	sklearn	-
	NumPy	1.19.2
	Pandas	1.1.2
	Scipy	1.5.2
	Keras ,Tensorflow, PyTorch, MxNet, GluonTS	-

3.3.2 Sandboxes

The following table provides information on the components that will be used within Pilot 5b sandbox. Table 9 - Pilot 5b Sandbox

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Transaction Categorization Engine	Analytics	DEP	N/A	Docker	NOT disclosed	Kubernetes	4
Cash Flow Prediction component	Analytics	DEP	N/A	Docker	NOT disclosed	Kubernetes	2
Budget Prediction engine	Analytics	DEP	N/A	N/A	N/A	N/A	N/A
KPI engine	Analytics	DEP	N/A	N/A	N/A	N/A	N/A
Transaction monitoring engine	Analytics	DEP	N/A	N/A	N/A	N/A	N/A
Invoice Processing engine	Analytics	DEP	N/A	N/A	N/A	N/A	N/A
Benchmark engine	Analytics	DEP	N/A	N/A	N/A	N/A	N/A
Recommender engine	Analytics	DEP	N/A	N/A	N/A	N/A	N/A

3.4 Pilot 6

Pilot 6 (“Personalized Closed-Loop Investment Portfolio Management for Retail Customers”) is managed by NBG. Large customer datasets and large volumes of customer-related alternative data sources (e.g., social media, news feeds, on-line information) will be used to feed ML/DL algorithms. The latter will provide the account officers with personalized, effective, and context-aware investment recommendations for the retail customers of the bank. The reader can find more information and a plan for Pilot 6 in Section 2.6 of D7.1.

3.4.1 Testbed Technical Specification

Pilot 6 testbed will be hosted on MS-Azure cloud infrastructure provided by NBG. The following table provides further details.

Table 10 - Pilot 6 Testbed

Testbed	Components
Hosting	Microsoft Azure
Server Infrastructure	3 VM instances, each with the following:VM1 -Standard B8ms Data Management Ubitech Icarus & Visualization Server - 8 vcpus, 32 GiB memory - 500GB HDDVM2 - Standard B4ms- Database Server(Leanxcale) - 4 vcpus, 16 GiB memory - 500GB HDDVM3 - Standard B4ms - Analytics Server- 4 vcpus, 16 GiB memory - 500GB
Managed Services	N/A
Other tools/ Applications	Reportbrain (RB) Sentiment Analysis API

3.4.2 Sandboxes

The following table provides information on the components that will be used within Pilot 6 sandbox.

Table 11 - Pilot 6 Sandbox

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
DataStore (Leanxcale)	Data Sources	legacy	4 vcpus, 16 GiB memory - 500GB HDD	Docker	NOT disclosed	Kubernetes	8
NBG Datasets	Data Sources	developed	N/A	none	NOT disclosed	N/A	
Data Collection (UBI Icarus)	Data Management	developed	8 vcpus, 32 GiB memory - 500GB HDD	Docker	NOT disclosed	Kubernetes	7
Data Normalization (UBI Icarus)	Security	developed	8 vcpus, 32 GiB memory - 500GB HDD	Docker	NOT disclosed	Kubernetes	7
Customer Risk Profile Cluster	Analytics	developed	4 vcpus, 16 GiB memory - 500GB HDD	none	NOT disclosed	Kubernetes	4
Personalized Investment Recommendation AI engine	Analytics	developed	4 vcpus, 16 GiB memory - 500GB HDD	none	NOT disclosed	Kubernetes	3
Customer initiation and personalized recommendation UI Application	Presentation	developed	8 vcpus, 32 GiB memory - 500GB HDD	Docker	Yes	Kubernetes	3

3.5 Pilot 7

Pilot 7 (“Operation Whitetail - Avoiding Financial Crime”) is managed by CXB. Due to a change in pilot partners, it is currently under development. The goal of Pilot 7 is to explore more accurate, comprehensive and near real-time pictures of suspicious behavior in Financial Crime, Fraud, and cyber-physical attacks having the final objective of stealing the bank customers’ identity and money. Bank internal as well as external data sources will be used to produce data giving insight to the financial crime risk score. This may include a risk score, customer data, transaction patterns and details. The reader can find more information for Pilot 7 in Section 2.7 of D7.1.

3.5.1 Testbed Technical Specification

The following table provides the information on the Pilot 7 testbed currently available. Table 12 - Pilot 7 Testbed

Testbed	Components
Hosting	The tesbed will be hosted on-premise for bank's own purposes and data privacy and in the INFINITECH AWS cloud utilizing the INFINITECH components
Server Infrastructure	The server infrastructure depends on the specific data volume, which needs to be defiend yet.The on-premise infrastructure includes a virtualized envrionment connected to the bank data storage
Managed Services	Virtualization, DBs, DNS
Other tools/ Applications	K8s

3.5.2 Sandboxes

The following table provides information on the components that will be used within Pilot 7 sandbox. Table 13 - Pilot 7 Sandbox

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Data pool	Data Source	Legacy	TBD		NOT Disclosed		
Data anonymization	Data Source	to be developed	TBD		NOT Disclosed		
Data Ingestion	Ingestion	to be developed	TBD		TBD		5
Data Extraction	Data Management	to be developed	TBD		TBD		5
Data Analytics / Scoring	Analytics	to be developed	TBD		TBD		4
Visualization	Presentation	to be developed	TBD		TBD		4

3.6 Pilot 8

Pilot 8 ("Platform for Anti Money Laundering Supervision (PAMLS)") is managed by BOS. The objective of the pilot is to develop a platform named PAMLS, namely Platform for anti-money laundering Supervision. PAMLS is supposed to improve the effectiveness of the existing supervisory activities in the area of anti-money laundering and combating financing of terrorism. To this aim, large quantity of data owned by BOS and other competent authorities (FIU) will be processed. The reader can find more information and a plan for Pilot 8 in Section 2.8 of D7.1.

3.6.1 Testbed Technical Specification

Testbed for Pilot 8 will be hosted on BOS premises. The following table provides further details. Table 14 - Pilot 8 Testbed

Testbed	Components
Hosting	On the premises of the BOS, it is ready and it has already deployed the software components and data to implement the PoC.
Server Infrastructure	Server: HP Z4 G4 WKS CPU: Intel XeonW-2125 4.0 4C RAM: 256GB (8x32GB) DDR4 Graphic: NVIDIA Quadro P400 2GB (3)mDP Graphics Storage: Z Turbo Drv 1TB PCIe NVMe OPAL2 TLC SSD
Managed Services	Windows Server 2019 Standard x64 operating system
Other tools/ Applications	N/A

3.6.2 Sandboxes

The following table provides information on the components that will be used within Pilot 8 sandbox.
Table 15 - Pilot 8 Sandbox

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Risk Calculation engine and Complex search services	analytics	to be developed		None	NOT disclosed	manual	TRL 5
Anomaly detection andprediction component	analytics	to be developed		Docker	Yes	manual	TRL 3
StreamStory component	analytics	to be developed		Docker	Yes	manual	TRL 4

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Pattern discovery and matchingcomponent	analytics	to be developed		Docker	Yes	manual	TRL 2
Pseudo-anonimization tool	Data Managenent	to be developed		Docker	Yes	manual	TRL 3
PostgreSQL	Data Managenent	Legacy	500 GB	None	open source	manual	TRL 9
ElasticSearch	Data Managenent	Legacy	200 GB	None	open source	manual	TRL 9
NEO4J	Presentation	Legacy	200 GB	None	open source	manual	TRL 9

3.7 Pilot 9

Pilot 9 ("Analysing Blockchain Transaction Graphs for Fraudulent Activities") is managed by AKTIF. It aims to leverage HPC technologies to analyze huge blockchain graphs, to detect fraudulent activities in crypto currencies transactions. The reader can find more information and a plan for Pilot 9 in Section 2.9 of D7.1.

3.7.1 Testbed Technical Specification

Testbed Pilot 9 is currently hosted on Amazon cloud. The following table provides further details. Table 16 - Pilot 9 Testbed

Testbed	Components
Hosting	Currently set up on Amazon AWS Hosting Partner Information Bogazici University and AktifBank Testbed Location info: Amazon cloud
Server Infrastructure	HPC Cluster on Amazon Cloud (16 c5.4xlarge instances), each instance having 16 virtual CPUs, 32 GiB memory and 500 GB SSD storage. A medium Amazon instance for running message queue.
Managed Services	N/A
Other tools/ Applications	Ubuntu Linux operating system
	StarCluster HPC cluster toolkit
	MPI message passing interface
	Rabbit MQ message queue
	Metis Parallel graph partitioner
	Vis.js open source graph visualization software for web interface

3.7.2 Sandboxes

The following table provides information on the components that will be used within Pilot 9 sandbox.
Table 17 - Pilot 9 Sandbox

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Blockchain Transaction Dataset Preparation Component	Data ingestion	developed	A medium Amazon instance	currently None/ docker planned	yes (https://zenodo.org/record/3669937#.YC-KzxMzY1I)	currently manual/ planned kubernetes	5

Component	RA	Type	Resources	Container	Repository	Deployment	TRL
Scalable Transaction Graph Analysis Component	Data Management / Analytics	developed	HPC Cluster on Amazon Cloud (16 c5.4xlarge instances), each instance having 16 virtual CPUs, 32 GiB memory and 500 GB SSD storage. A medium Amazon instance for running message queue.	currently None/ docker planned		currently manual/ planned kubernetes	5
User Interface for Blockchain Transaction Reports and Visualization Component	Interface / Analytics	developed	A large Amazon instance	None		manual	3

3.8 Pilot 10

Pilot 10 ("Real-time cybersecurity analytics on financial transactions' data") is managed by PI. It aims to speed up the detection of suspected fraudulent transactions and to identify security-related anomalies while they are occurring. This objective can be pursued through the real-time analysis of the financial transactions of home and mobile banking systems. The reader can find more information and a plan for Pilot 10 in Section 2.10 of D7.1.

3.8.1 Testbed Technical Specification

Testbed for Pilot 10 will be hosted on AWS environment. The following table provides further details.

Table 18 - Pilot 10 testbed

Testbed	Components
Hosting	GFT's AWS environment
Server Infrastructure	On demand over EKS
Storage	TBD
Managed Services	N/A
Other tools/Applications	Tools available in the DevOps Sandbox

3.8.2 Sandboxes

The following table provides information on the components that will be used within Pilot 10 sandbox.

Table 19 - Pilot 10 Sandbox 1

Components	Service	Type	Resources	Container	Repository	Deployment	TRL
Gitlab Community Edition	Source code management	legacy	TBD	Docker	TBD	v11.7.5	9
Docker	Container	legacy	TBD	Docker	TBD	v1.13.1, build: 7f2769b/1.13.1	9
Kubernetes	Orchestrator	legacy	TBD	Docker	TBD	v1.11.0+d4cacc0.	9
OKD it is an opensource layer on top of Kubernetes.	Orchestrator	legacy	TBD	Docker	TBD	v3.11.0+ec8630f-265,	9
OpenShift Container Registry storage of OKD docker images	Registry	legacy	TBD	Docker	TBD	v3.11.0+ea42280	9
Nexus Sonatype : storage of services/ application docker images and DL models binaries and artefacts.	Registry	legacy	TBD	Docker	TBD	v3.20.1-01, OSS Edition	9
Jenkins	CI/CD software	legacy	TBD	Docker	TBD	v2.204.1	9

Table 20 - Pilot 10 Sandbox 2

Components	RA	Type	Res.	Container	Repository	Deployment	TRL
Visualization tool (T4.6)	Interface	Infinitech	TBD	docker	Docker Registry hosted in the DevOps Sandbox	kubernetes	N.A.
Pseudo-anonymization tool (T3.5)	Data Management	Infinitech	TBD	docker	Docker Registry hosted in the DevOps Sandbox	kubernetes	N.A.
Transaction Generator	Data Source	Infinitech	TBD	docker	Docker Registry hosted in the DevOps Sandbox	legacy	N.A.
Alida	Analytics and Machine Learning	Developed	4 cpu, 8gb RAM	docker	not disclosed	kubernetes (sidecar deployment, see D6.10)	6

3.9 Pilot 15

Pilot 15 (“Inter-Banking Open Pilot”) is managed by ABILAB. It aims to leverage Machine Learning and Natural Language Understanding paradigms to implement the prototype of a solution that could address and tackle shared business pains among several banks. The solution will read and analyze extensive internal documentation of banks in real time to highlight the main concepts and compare them with reference taxonomies to build a common business glossary. Pilot 15 will be hosted and deployed on the Testbed blueprint that will be developed accordingly to the pilot requirements. The reader can find more information on Pilot 15 in Section 2.15 of D7.1.

3.9.1 Testbed Technical Specification

Pilot 15 testbed finds a description in the reference (Blueprint) testbed of the INFINITECH project alredy provided in Section 2.1 and reported also in the following table for convenience.

Table 21 - Pilot 15 Testbed

Testbed	Components
Hosting/Cloud	AWS
Server Infrastructure	EC2 Instances
Managed Services	Kubernetes EKS

3.9.2 Sandboxes

Pilot 15 sandboxes are still to be defined and will be reported in the following versions of the document.

4 Conclusions and Future Work

This deliverable has introduced the first version of the Testbed and Sandboxes of the incumbent operators and should be considered as a first assessment of the project's Pilots. Albeit it contains the information gathered by the Pilots' Partners with the description and definition of each Testbed and Sandbox components in an advanced stage, it will be completed in other two iterations and the outcomes will be presented in successive versions of this deliverable.

Whether the information contained into this document is to be considered as list of the resources needed to perform the demonstrators is up to the organization that manages the actual infrastructure: in fact it can actually be used as input to IT operators and Financial cost structures to set up the testbeds and therefore it is extremely valuable to organizations from IT to Procurement departments.

As an important outcome of the Task activities on T6.4 and this assessment, the deliverable presents the information about 8 testbeds instantiated and one more as a result of the Reference Infrastructure (Pilot 15). This is in line with the first KPI of Objective 6 in WP6 that can be considered achieved with 8 testbeds instantiated and one more as a result of the Reference Infrastructure (Pilot 15).

As per the specific KPIs set for the project the following table addresses the specific indexes listed in the DoA.

Table 22 - Mapping of INFINITECH DoA/Task KPI with Deliverable Achievements

KPI	Description	Comment
KPI#6.1	Testbeds to be Established >= 9 (>=8 in Incumbent organizations and >=1 (EU-wide) testbed for FinTech/InsuranceTech firms);	This Deliverable presents 8 testbeds actually managed in 8 different pilots. The reference infrastructure hosts the Pilot 15 that is to be considered the 9th (i.e. 8+1).

This deliverable is the first of other two versions namely Deliverable D6.8 an D6.9, that will follow in September 2021 and in June 2022 and will complete the information of physical testbeds and logical sandboxes with more detailed information.

5 References

[1] Kubernetes IP management https://kubernetes.io/docs/concepts/services-networking/service/#choosing-your-own-ip-address

[2] Kubernetes. [Online] https://kubernetes.io

[3] «Hubble,» [Online]. Available: https://docs.cilium.io/en/v1.9/intro/#intro.

[4] VMware. [Online]. Available: https://www.vmware.com/.

[5] v. ESXi. [Online]. Available: https://www.vmware.com/products/esxi-and-esx.html.

[6] Vagrant. [Online]. Available: https://www.vagrantup.com/.

[7] «Terraform,» [Online]. Available: https://www.terraform.io/.

[8] «Terraform homepage,» HashiCorp, [Online]. Available: https://www.terraform.io/docs/language/files/index.html.

[9] «Oralce Virtual Box,» [Online]. Available: https://www.virtualbox.org/.

[10] «VMware Workstation,» [Online]. Available: https://www.vmware.com/products/workstation-pro.html.

[11] Ruby. [Online]. Available: https://www.ruby-lang.org/en/.

[12] «EC",» [Online]. Available: https://aws.amazon.com/ec2/instance-types/.

[13] «HCL,» [Online]. Available: https://www.terraform.io/docs/configuration-0-11/syntax.html.

[14] «Cloning,» [Online]. Available: https://sdorsett.github.io/post/2018-12-24-using-terraform-to-clone-a-virtual-machine-on-vsphere/).

[15] «rancher,» [Online]. Available: https://rancher.com/.

[16] «CNCF,» Linux Foundation, [Online]. Available: https://www.cncf.io/.

[17] «Open-LDAP,» Openldap Foundation, [Online]. Available: https://www.openldap.org/.

[18] «MetalLB,» [Online]. Available: https://metallb.org/.

[19] «HELM,» [Online]. Available: https://helm.sh/.

[20] «GitLab homepage,» GitLab, [Online]. Available: https://about.gitlab.com/.

[21] «Harbor homepage,» Linux Foundation, [Online]. Available: https://goharbor.io/.

[22] «Jenkins homepage,» [Online]. Available: https://www.jenkins.io/.