CEF Digital
Connecting Europe

BDTI Technical Webinar

Walkthrough Identification of Humanitarian need use case – Using BDTI services

DIGIT
Directorate-General
for Informatics

DG Connect
Directorate-General for Communications Networks,
Content and Technology

Instructions for the live webinar:

This is an interactive session. There will be time for questions and answer throughout the presentation via Slido. We hope you will share your views.

Click on *Connect audio* to hear the presenters but please mute your microphones when you are not speaking.

Please note that this webinar will be recorded.

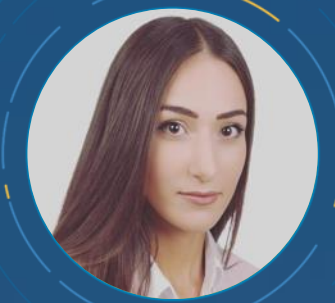
CEF Digital
Connecting Europe

Welcome to the BDTI Technical Webinar

Agenda for today

1. Introduction to BDTI (10 min)
2. BDTI Architectural Service Offerings (10 min)
3. Use case: Identification of humanitarian need (5 min)
4. Selecting BDTI services for the use case (5 min)
5. Accessing the environment and importing the data (15 min)
6. Using BDTI Self-Paced Labs (25 min)
7. Support & How to apply (10 min)
8. Q&As (10 min)

CEF Digital
Connecting Europe



Francesca Vella
Pilot Request Manager
BDTI Team

Introduction to BDTI (10 min)

What is the Big Data Test Infrastructure

*The **Big Data Test Infrastructure** will provide a set of **data and analytics services**, from infrastructure, tools and stakeholder onboarding services, allowing European public organisations to **experiment with Big Data technologies** and move towards **data-driven decision making***

BDTI Initiative drivers

Problem - *Solution*

Lack of Big Data technologies

Facilitate the prototyping and launching of pilot

Lack of Big Data skills

Facilitate Big Data knowledge in public sector

Data sharing among public organisations is not yet a common practice

Provide built-in connectors/APIs and foster the sharing of data sources to better support policy-making

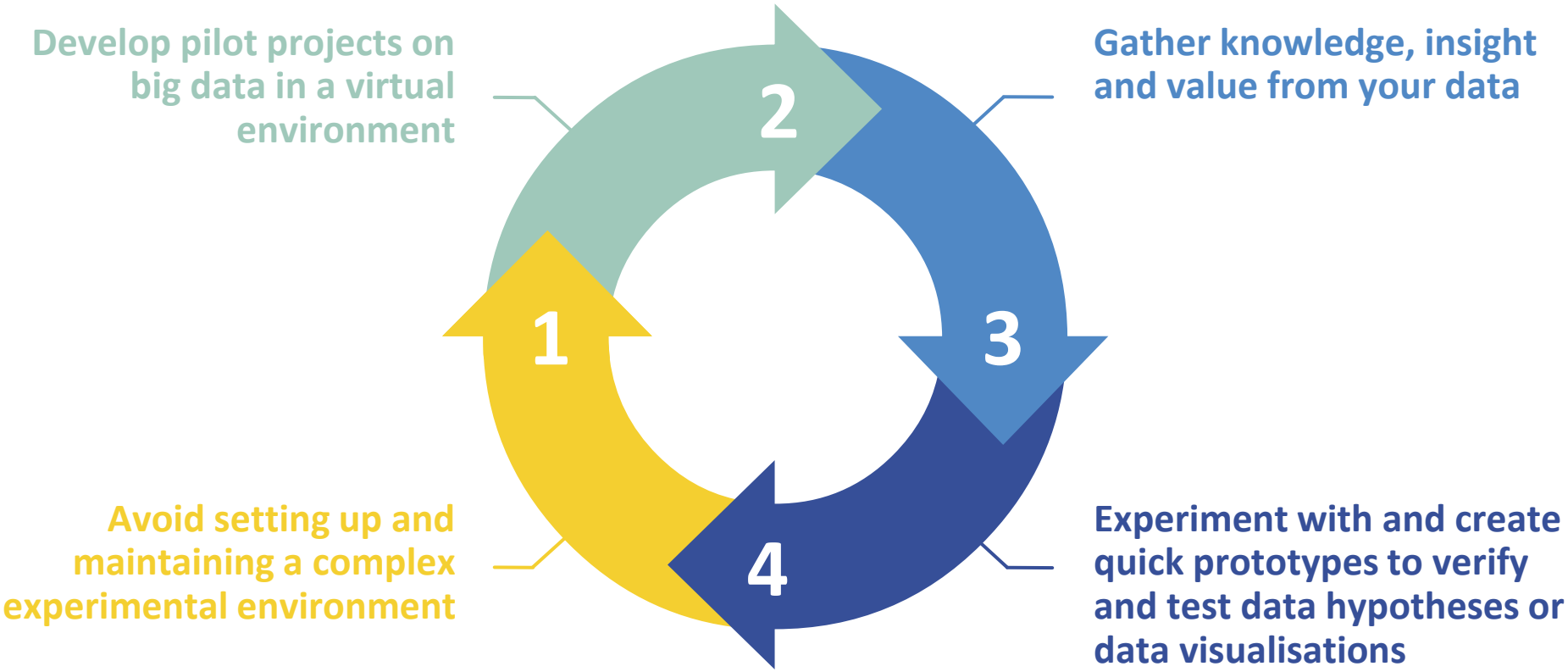
Risk of replicating the efforts by implementing similar projects

Support public organisations through the creation of a Big Data community for the sharing of good practices, pilot outcomes, etc.

Is BDTI for me?

Yes, if you need to experiment with big data in a safe environment.

What can we help you achieve?



BDTI Use Cases (1/2)

Descriptive Analysis

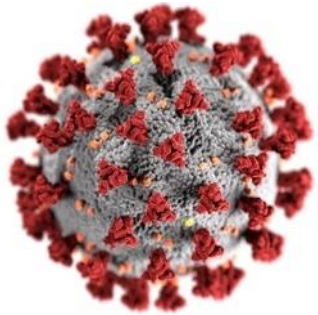
Use of statistics to quantitatively describe features of a collection of information

Social Media Analysis

Gather and analyse data from social media to improve business decisions

Time-series Analysis

Analyse time series data in order to extract meaningful statistics and other data characteristics



Covid 19 Analysis

Analyse Covid 19 information extracted from a dataset collecting data taken from medical papers

Predictive Analysis

Use statistical techniques that analyse current and historical facts to make predictions about future or unknown events

Network Analysis

Investigate any structures through the use of network and graph theories

Text Analysis

Use natural language processing to analyse unstructured text data, to derive pattern and trends

BDTI Use Cases (2/2)

Web Analysis (Scraping/Monitoring)
Gathering information from websites, involving data scraping (using both or web-crawler) and data parsing to extract un-organised web data, as well as data from API's, into manageable format.

Population/Customer segmentation
Dividing a broad population into sub-groups of consumers based on shared characteristics such as common needs, interests, similar lifestyles or even similar demographic profiles.

Route-Traceability/Flow monitoring
Tracking and detection of objects through the use of sensors (e.g. GPS, mobile phone signals, road cameras) or any other types of data usable for this purpose.

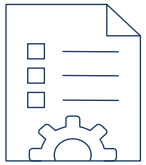
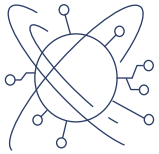
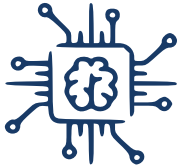


Image Processing
Computational operations using any form of signal processing for which the input is an image, a series of images, or frames of a video; output of image processing may be either an image or a set of characteristics / parameters related to the image.



IoT & Smart City
Gathering relevant information in IoT environments, in a Smart City context.

IoT Security
Safeguarding connected IoT devices and networks, since security often has not been considered in IoT product design.



Applying Bioinformatics to genetic data
The use of computational biology, applying data science techniques to understanding/organising biological information and analysing genetic data

Main Benefits

Interoperable environments and tools that use open source technologies, ensuring their integration with other systems

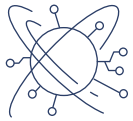
High **performance** due to an environment architecture that easily scales resources needed for dealing with big data.

Scalability due to an environment architecture tailored to the required storage size and computing resources

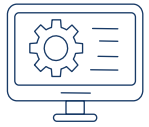
Reliability and **availability** during data transfer and data storage

Modifiability for using big data evolving tools and technologies

Necessary **security** implementations for safe data experimentation



Share and re-use data across policy domains and organisations



Access to a **knowledge base** and **advisory** for the implementation of pilot project

Access to **insights on best practices** with big data projects and other pilots

Implement a **free of charge** pilot project

CEF Digital
Connecting Europe

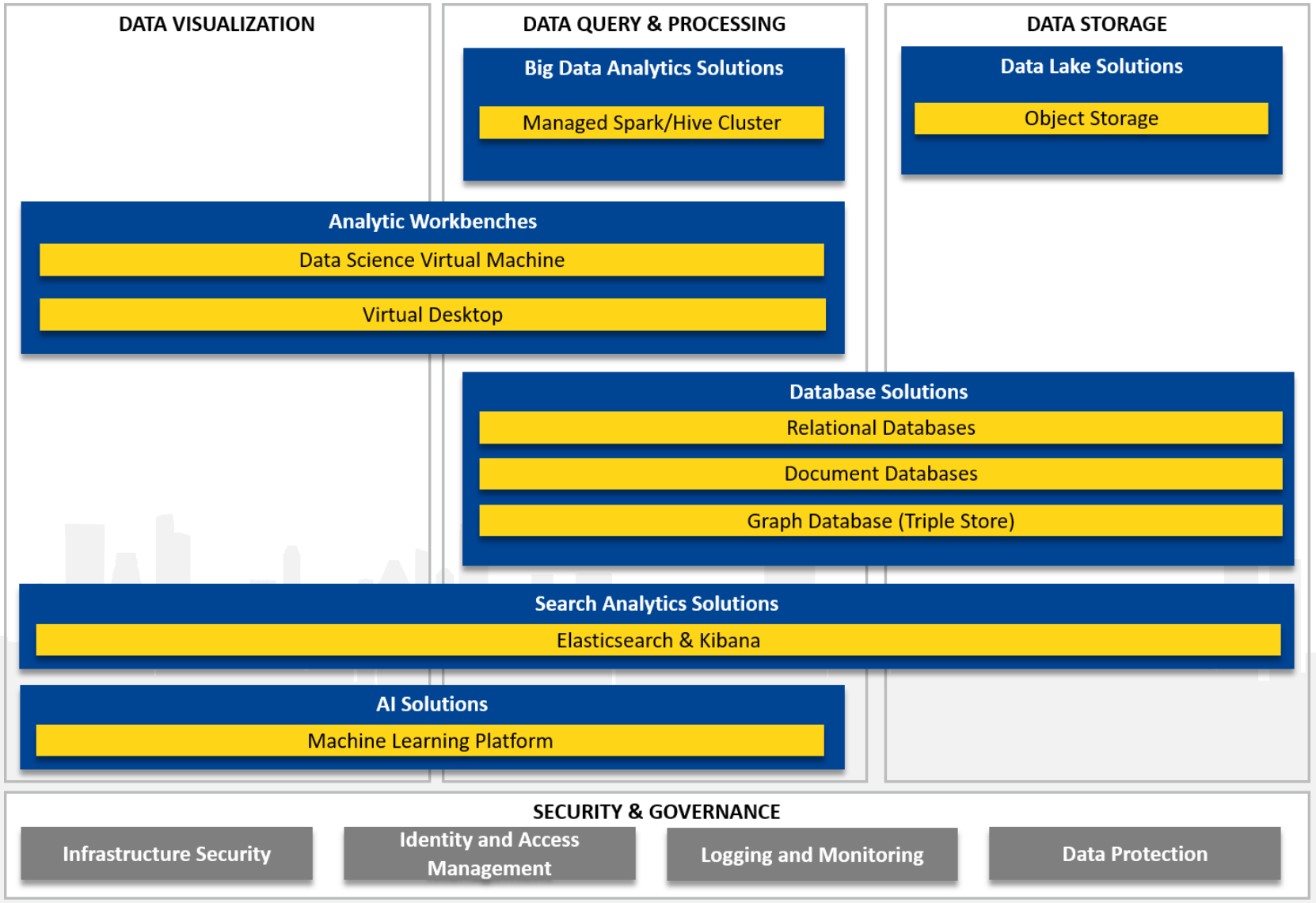
BDTI Architectural Service Offerings (10 min)

Kasper Rutten

Cloud Engineer

BDTI Team

General BDTI service offering



Analytic Workbenches

An analytics workbench empowers users with the ability to autonomously produce and publish insights, mainly through self-service data preparation and visual data discovery tools.

WorkSpaces

Connect to a full-featured Windows Desktop environment through a remote client. Jupyter notebooks, Rstudio, KNIME Analytics, Anaconda are available and can be used for processing, analyzing and visualizing data.

- A Virtual Desktop as your Analytics Workbench, perform ad-hoc analysis with Jupyter notebooks, KNIME Analytics or RStudio.
- Access data from Database & Data Lake solutions
- Store processed data from Database & Data Lake solutions
- Visualize analysis with Jupyter, Rstudio or your favorite plotting library.

Data Science Virtual Machine

Connect to a full-featured Ubuntu 18.04 LTS Desktop environment through your browser. Jupyter notebooks, RStudio Server or a Terminal session can be accessed securely from your browser.

- A DSVM as your Analytics Workbench, perform ad-hoc analysis with Jupyter notebooks or RStudio.
- Access data from Database & Data Lake solutions
- Store processed data from Database & Data Lake solutions
- Visualize analysis with Jupyter or Rstudio
- Execute jobs on Spark/Hive cluster with Jupyter sparkmagic kernels

Data Lake Solutions

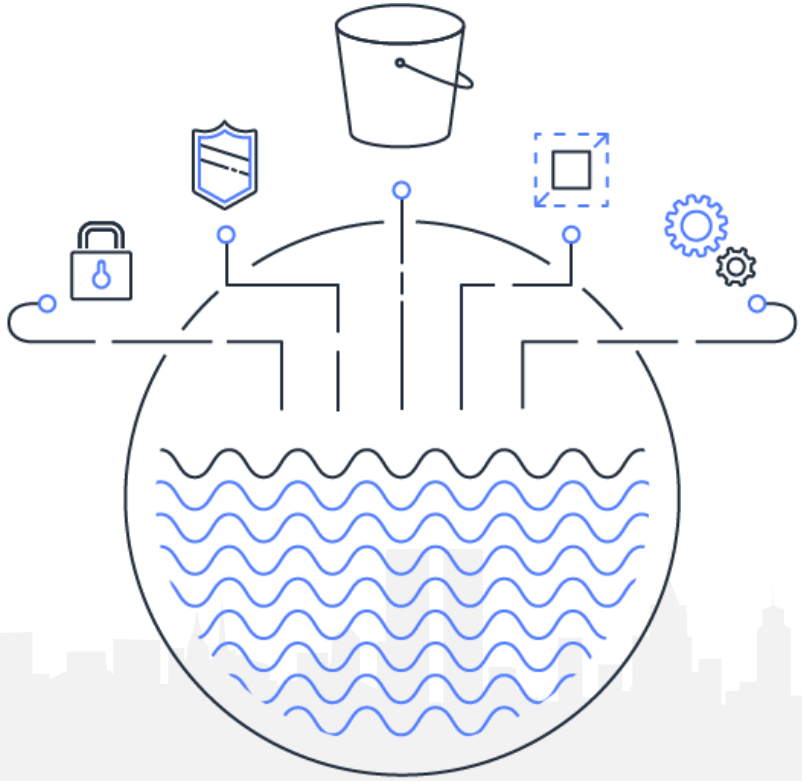
A Data Lake is a repository of data stored in its raw format. It is the main starting place for self-service analytics.

Typically, Data Lakes store large volumes of information, classified as Big Data.

Object Storage

Amazon S3 is an object storage solution that stores objects up to 5 TB in buckets. In your pilot you can upload data to your S3 bucket(s) and structure them accordingly in folders and subfolders. You can upload as many objects as you want in Amazon S3, storage capacity will automatically scale.

- Use object storage as your primary storage for raw data (data lake solution)
- Starting point for big data analytics
- Access object storage from other BDTI resources such as a Data Science Virtual Machine or Spark/Hive Cluster



Big Data Analytics Solutions

Analyzing large data sets requires significant compute capacity that can vary in size based on the amount of input data and the type of analysis.

Specialized tooling is therefore also necessary.

Managed Spark/Hive Cluster

Using open-source tools such as Apache Spark and Apache Hive coupled with Object Storage enables performing big data analytics on vast amounts of data.

A Spark/Hive cluster is implemented with Amazon EMR which allows leveraing EMRFS, an S3 interface for Hadoop workloads.

- Perform big data analytics on large amounts of data
- Leverages the Hadoop ecosystem, an open-source framework for distributed processing and big data analytics
- Execute Spark and Hive jobs on HDFS or Object Storage (Amazon S3)
- Leverage the Spark API for advanced analytics and processing
- Leverage Hive for Big Data warehousing

Database Solutions

Database systems are the most well-known and standardized solutions for data storage and querying. Different flavors are available depending on the use case.

Relational Database

Uses open-source relational databases such as MySQL and Postgres.

Relational databases are implemented with Amazon RDS.

- Perfect fit for most transactional and analytical processing (OLTP & OLAP cases)
- Perform flexible SQL queries to extract and store data
- Can be scaled depending on the needs

Document Store

Leverage the NoSQL paradigm to store data as a collection of documents to allow further scaling of your data storage solution.

Document stores are implemented with Amazon DocumentDB.

- NoSQL solution, does not comply with the ACID constraints to allow better performance and scaling for specific use cases.
- Stores data as a colletion of documents.
- Query data with a Mongo-compatible API.

Search & Analytics Solutions

Search & Analytics solutions provide users with the ability to perform search queries on large amounts of text documents and in addition analyze its content.

Elasticsearch & Kibana

Use an open-source search engine to perform text search and analytics on large collections of documents. Visualize insights with Kibana.

Elasticsearch & Kibana is implemented with Amazon Elasticsearch Service

- Search and analytics engine for performing real-time search in text-based data.
- Graphical user interface for visualization and querying (Kibana)
- Operational analytics (log analysis).

AI Solutions

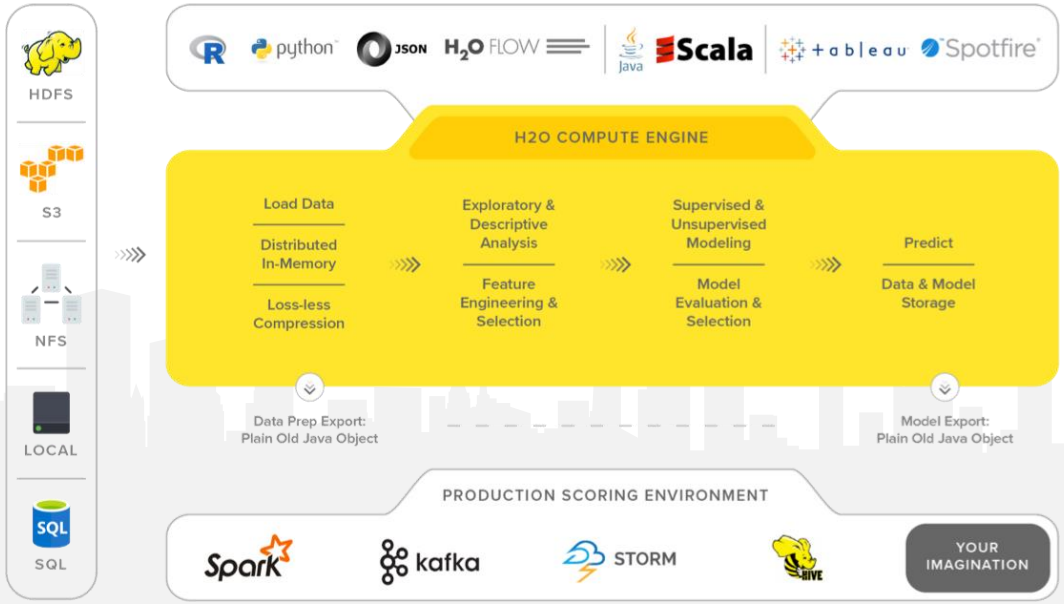
AI Solutions are specialized tools and frameworks to provide users with the ability to develop and publish artificial intelligence in the form of Natural Language Processing, Machine Learning, Data Mining and Predictive Modeling.

Machine Learning Platform (h2o.ai)

Use an easy-to-use open-source platform to train machine learning models on your data.

Machine Learning Platform is implemented with h2o.ai

- Supports the most widely used statistical & machine learning algorithms (including gradient boosted machines, generalized linear models, deep learning and more).
- Use H2O Flow, an open-source user interface for H2O. It is a web-based interactive environment that allows you to combine code execution, text, mathematics, plots, and rich media in a single document



BDTI Security & Governance

Infrastructure Security

BDTI takes care of all operational aspects regarding infrastructure security (network security, disaster recovery, maintenance, …)

Identity & Access Management

BDTI offers a managed centralized solution for identity and access management. Users can access all resources of their pilot through one BDTI account.

Logging & Monitoring

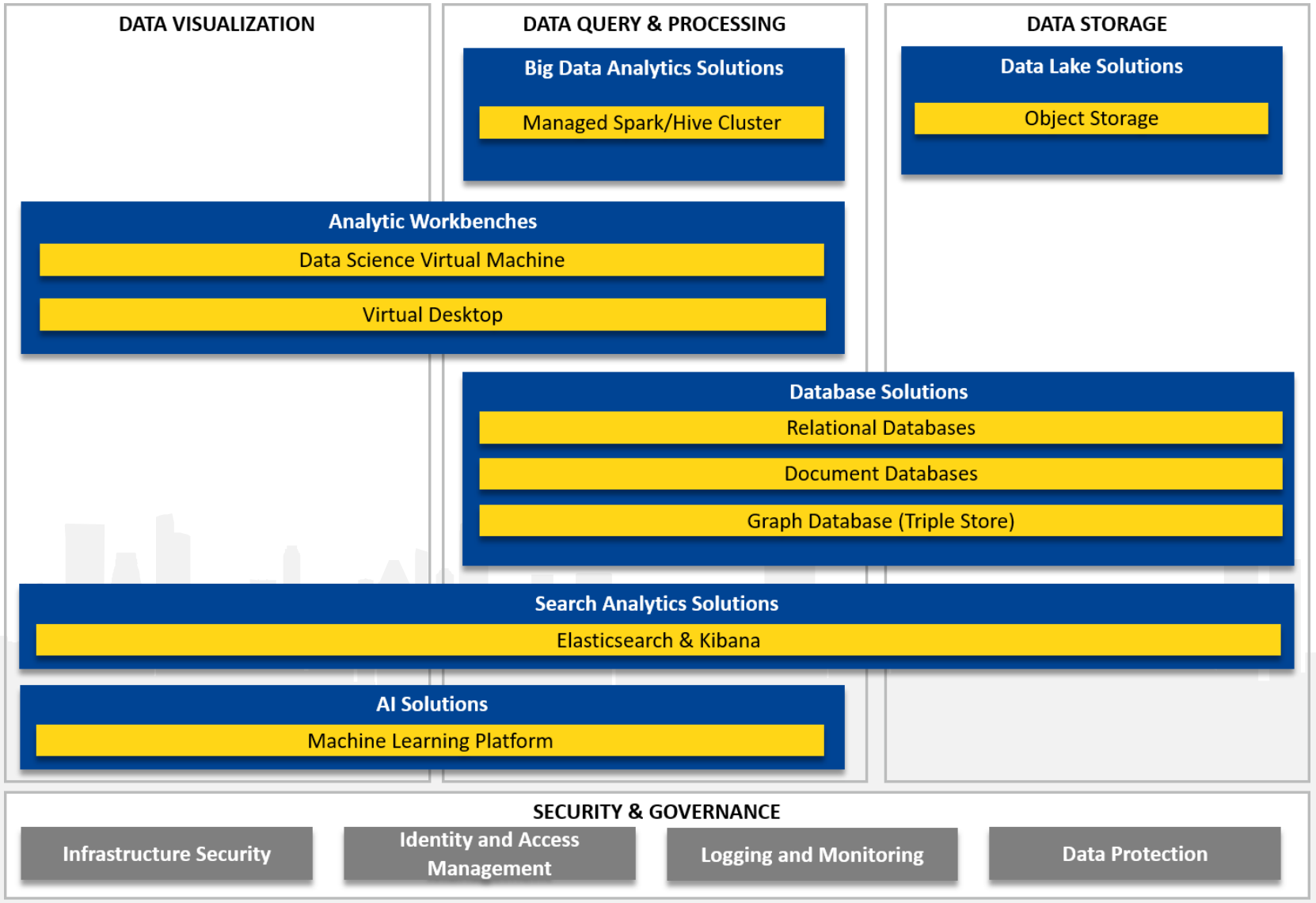
BDTI monitors all pilots for incidents and has logging in place for auditing and traceability.

Data Protection

BDTI protects all data that is being stored and processed in pilot environments.

All access needs to be authenticated and authorized. Data is encrypted in-transit and at-rest.

General BDTI service offering



CEF Digital
Connecting Europe

Use case: Identification of humanitarian need (5 min)

Nicolas Cattoir

Cloud Solution Architect

BDTI Team

Use case description

Situation

HELP International is an **international humanitarian NGO** that is committed to **fighting poverty** and providing the people of under-developed countries with **basic amenities and relief**.
After a fundraising request, HELP International has been able to **raise almost € 10 million.**

Problem

The CEO of the NGO needs to decide how to use this money strategically and effectively. So, the CEO has to make decision to **choose the countries that are in the direst need of aid**. Hence, your Job as a Data scientist is to **cluster the countries using some socio-economic and health factors** that determine the overall development of the country. Based on this information, the CEO should gain better insights into the next steps on how to allocate the money.

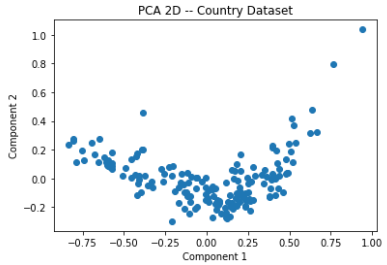
Use case description

Code snippets available in Self-paced labs

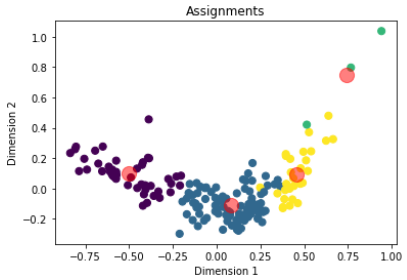
Approach

Country Dataset (Kaggle)

Principal Component Analysis (PCA)



Clustering



- ***child_mort***: Death of children under 5 years of age per 1000 live births.
- ***exports***: Exports of goods and services per capita. Given as %age of the GDP per capita.
- ***health***: Total health spending per capita. Given as %age of GDP per capita.
- ***imports***: Imports of goods and services per capita. Given as %age of the GDP per capita.
- ***income***: Net income per person.
- ***inflation***: The measurement of the annual growth rate of the Total GDP.
- ***life_expec***: The average number of years a newborn child would live if the current mortality patterns are to remain the same.
- ***total_fer***: The number of children that would be born to each woman if the current age-fertility rates remain the same.
- ***gdpp***: The GDP per capita. Calculated as the Total GDP divided by the total population.

167 rows x 10 columns

CEF Digital
Connecting Europe

Selecting BDTI services for the use case (5 min)

Nicolas Cattoir

Cloud Solution Architect

BDTI Team

Overview of your BDTI infrastructure

1

Data Science Studio

A Data Science Virtual Machine is a Linux-based virtual machine with common pre-installed tools and libraries for data scientists. This building block should be used for hosting collaborative development environments.

2

Object Storage

Highly durable, available, and performant object storage for frequently accessed data. Allows sharing object storage with multiple users. A perfect fit for big data analytics.
Implemented with Amazon S3.

3

Virtual Desktops

Virtual Desktops is a managed, secure Desktop-as-a-Service (DaaS) solution. Securely access common data science tools such as Jupyter, Rstudio, Anaconda or a Terminal session directly in your personal virtual desktop.
Implemented with Amazon WorkSpaces.

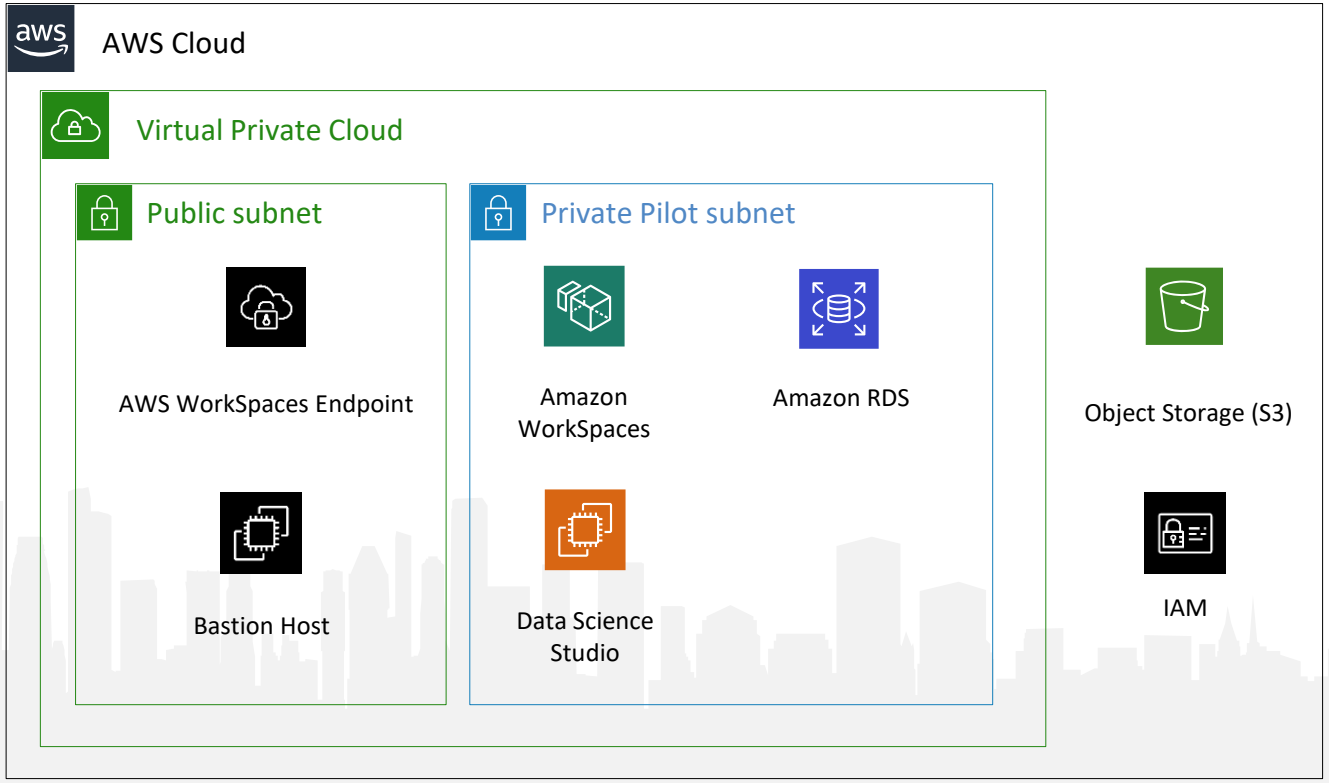
4

Relational Database storage

Production-ready relational database to store structured data with no software installation required.
Implemented with Amazon RDS.

Your BDTI infrastructure

What is the final networking of your BDTI services?



CEF Digital
Connecting Europe

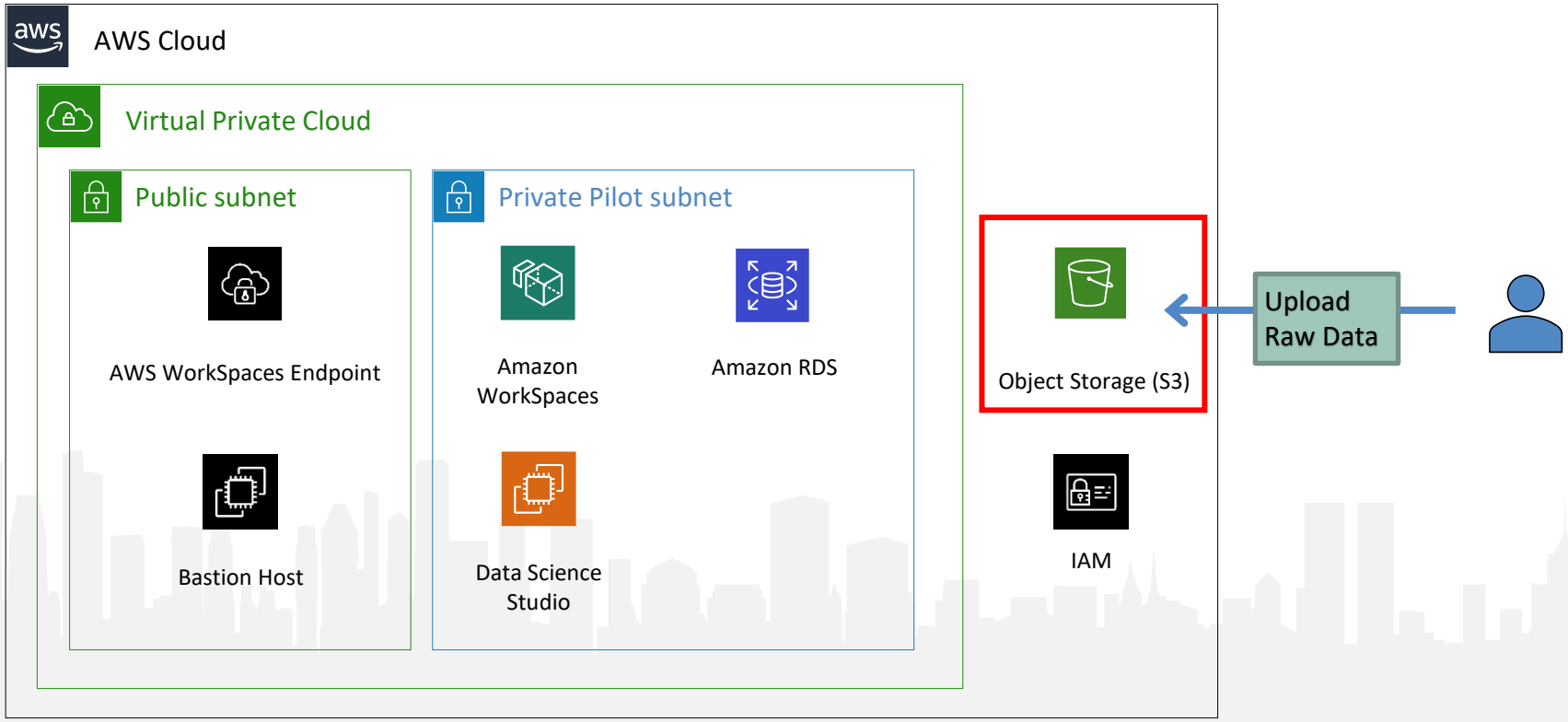
Accessing the environment and importing the data (15 min)

Ze Wen Wu

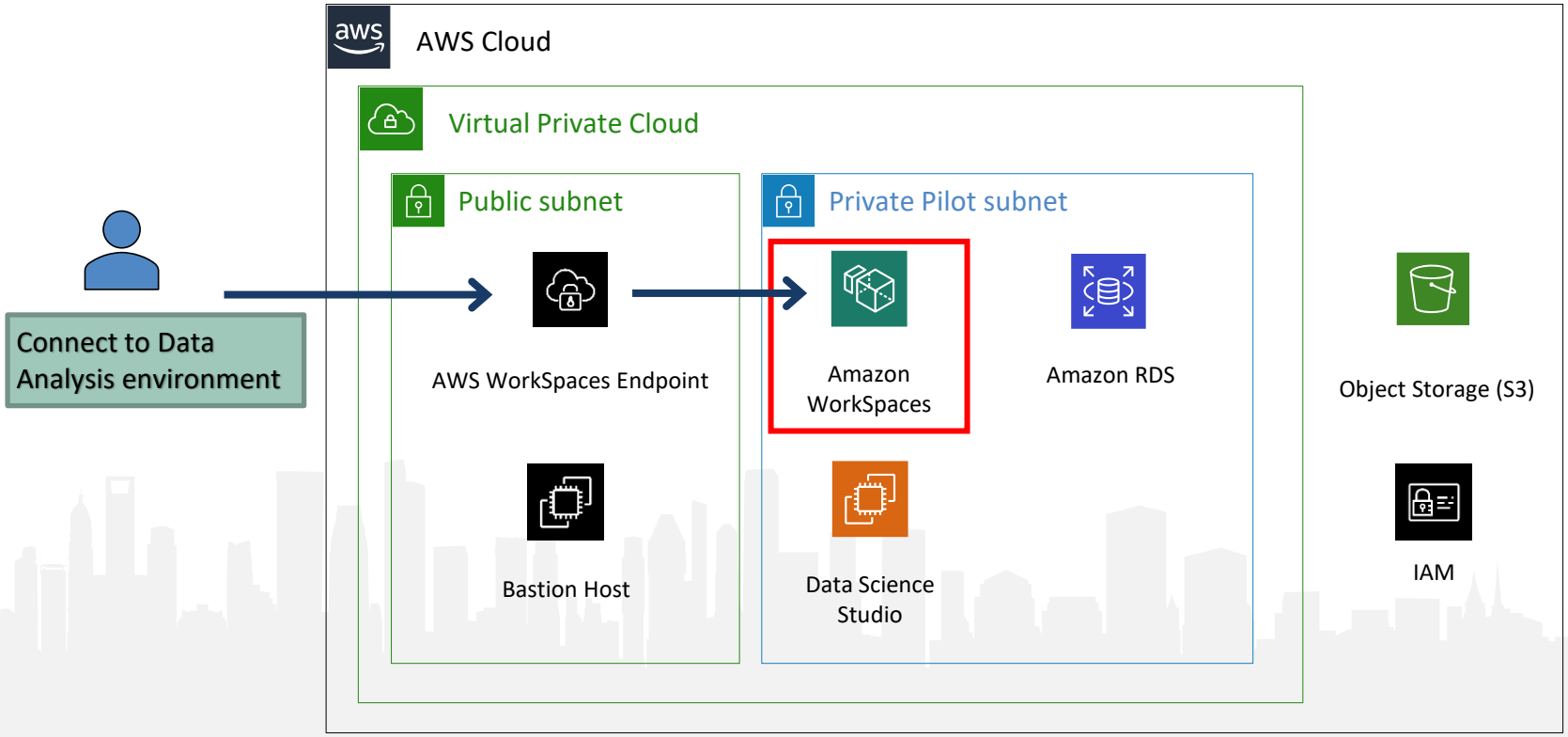
Cloud Engineer

BDTI Team

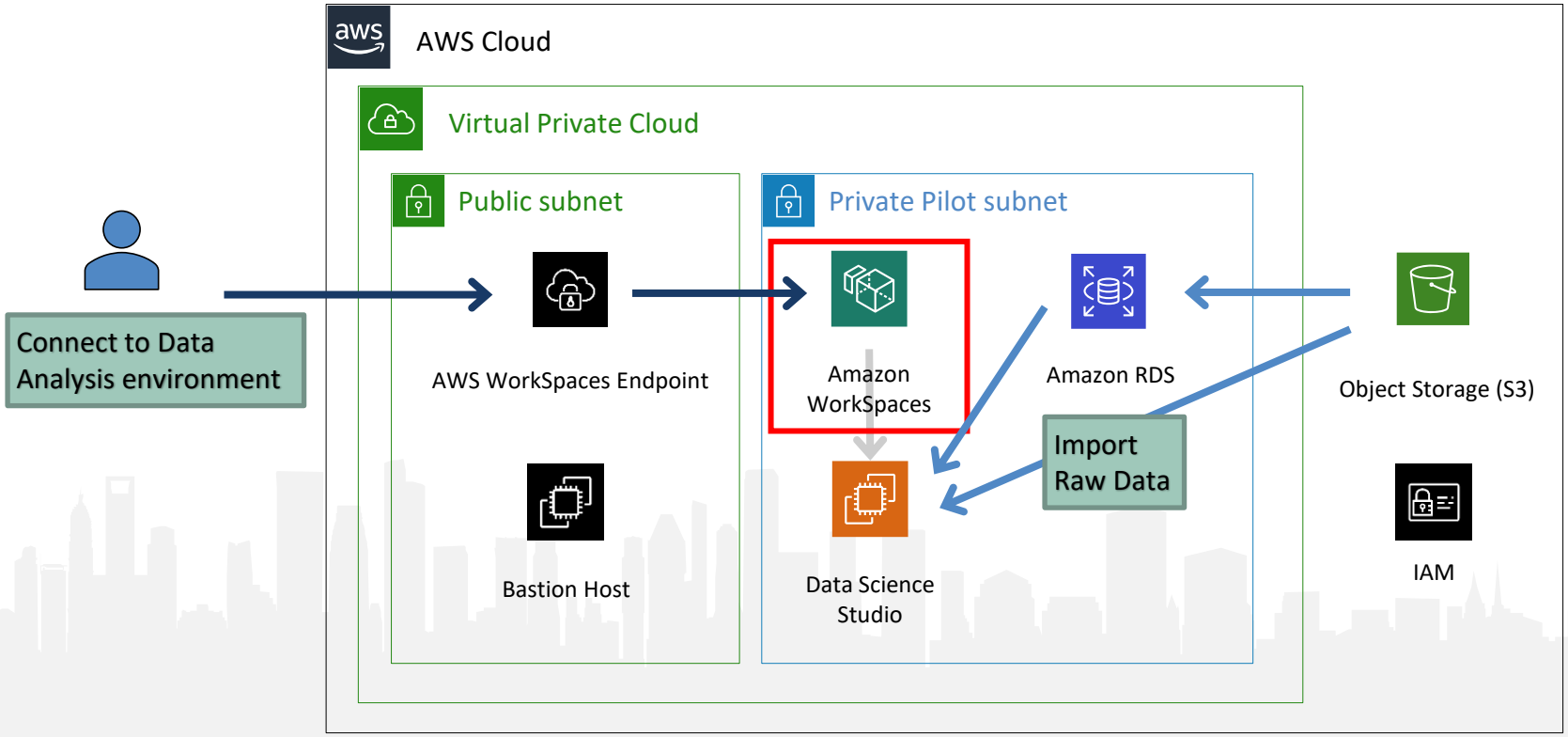
Connection to BDTI Infrastructure – Import Raw Data into S3



Connection to BDTI Infrastructure – Access WorkSpaces to analyze Raw Data



Connection to BDTI Infrastructure – Access WorkSpaces to analyze Raw Data



CEF Digital
Connecting Europe

Using BDTI Self-Paced Labs (25 min)

Ze Wen Wu

Cloud Engineer

BDTI Team

What are self-paced labs?

Here is an overview of what self-paced labs are.

Jupyter Notebook

- Text and code integration for **interactive learning experience.**

Frameworks

- Introduction popular **Python frameworks** for working with data.

Open source data

- Using data from **EU Open Data Portal** and **kaggle**.

What are the currently available self-paced labs?

Here are some of the self-paced labs we have developed.

LAB	DATA	TOOLS
Onboarding: Local Computer	/	• Conda • Boto3
Onboarding: Data Science Studio	/	• Conda • Boto3 • Mysql.connector • elasticsearch
Python: Data Exploration	**EU Open Data:** Belgium COVID-19	• Pandas DataFrame • Matplotlib
Machine Learning: Regression	**EU Open Data:** Gender Equality Index Dataset	• Scikit-learn
Machine Learning: Classification	**Kaggle:** Breast Cancer Wisconsin Dataset	• Scikit-learn
Machine Learning: Clustering	**Kaggle:** Country Dataset	• Scikit-learn
Apache Spark: Data Exploration	**EU Open Data:** Belgium COVID-19	• PySpark DataFrame • Matplotlib
Apache Hive: Data Exploration	**EU Open Data:** Belgium COVID-19	• HiveQL • Matplotlib

Process of Data analytics

What are the steps of data analytics?

Process of Data analytics

What are the steps of data analytics?

Process of Data analytics – Defining goals & Preprocessing

Goal:

- Identify the countries who can benefit the most from financial aid, using the following data from countries:

- ***child_mort***: Death of children under 5 years of age per 1000 live births.
- ***exports***: Exports of goods and services per capita. Given as %age of the GDP per capita.
- ***health***: Total health spending per capita. Given as %age of GDP per capita.
- ***imports***: Imports of goods and services per capita. Given as %age of the GDP per capita.
- ***income***: Net income per person.
- ***inflation***: The measurement of the annual growth rate of the Total GDP.
- ***life_expec***: The average number of years a newborn child would live if the current mortality patterns are to remain the same.
- ***total_fer***: The number of children that would be born to each woman if the current age-fertility rates remain the same.
- ***gdpp***: The GDP per capita. Calculated as the Total GDP divided by the total population.

Process of Data analytics – Defining goals & Preprocessing

Preprocessing:

- **Detect and fix issues in the data**
  (such as dealing with missing values, errors, or outlier values),
- **Deriving new feature columns** by transforming or combining existing features
  (a process known as feature engineering),
- **Normalizing numeric features** (values you can measure or count) so they're on a similar scale,
- **Encoding categorical features** (values that represent discrete categories) as numeric indicators.
- …

Process of Data analytics – Defining goals & Preprocessing

Preprocessing:

- ~~**Detect and fix issues in the data**~~
  ~~(such as dealing with missing values, errors, or outlier values),~~

- ~~**Deriving new feature columns** by transforming or combining existing features~~
  ~~(a process known as feature engineering),~~

- **Normalizing numeric features** (values you can measure or count) so they're on a similar scale,

- ~~**Encoding categorical features** (values that represent discrete categories) as numeric indicators.~~

- ***child_mort***: Death of children under 5 years of age per 1000 live births.
- ***exports***: Exports of goods and services per capita. Given as %age of the GDP per capita.
- ***health***: Total health spending per capita. Given as %age of GDP per capita.
- ***imports***: Imports of goods and services per capita. Given as %age of the GDP per capita.
- ***income***: Net income per person.
- ***inflation***: The measurement of the annual growth rate of the Total GDP.
- ***life_expec***: The average number of years a newborn child would live if the current mortality patterns are to remain the same.
- ***total_fer***: The number of children that would be born to each woman if the current age-fertility rates remain the same.
- ***gdpp***: The GDP per capita. Calculated as the Total GDP divided by the total population.

Preprocessing:

- **Normalizing numeric features** (values you can measure or count) so they're on a similar scale

Removing unit dependency using, e.g., **MinMax scaler:**

$$0 \leq x_i \leq 1 \text{ for all features } x_i$$

$$x_i = \frac{x_{old,i} - \min_i x_{old,i}}{\max_i x_{old,i} - \min_i x_{old,i}}$$

Basic idea: All numerical features should be in approximately same range.
Alternatives: Clipping, Log scaling, Z-score scaling, …

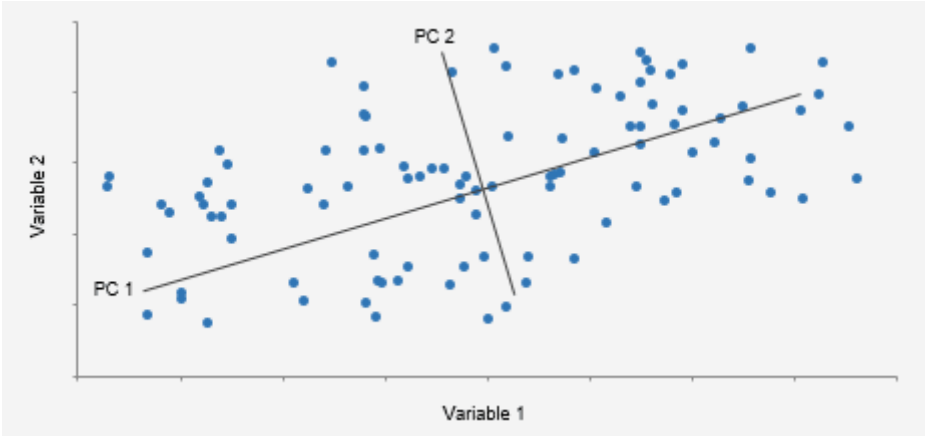
Process of Data analytics

What are the steps of data analytics?

Visualization:

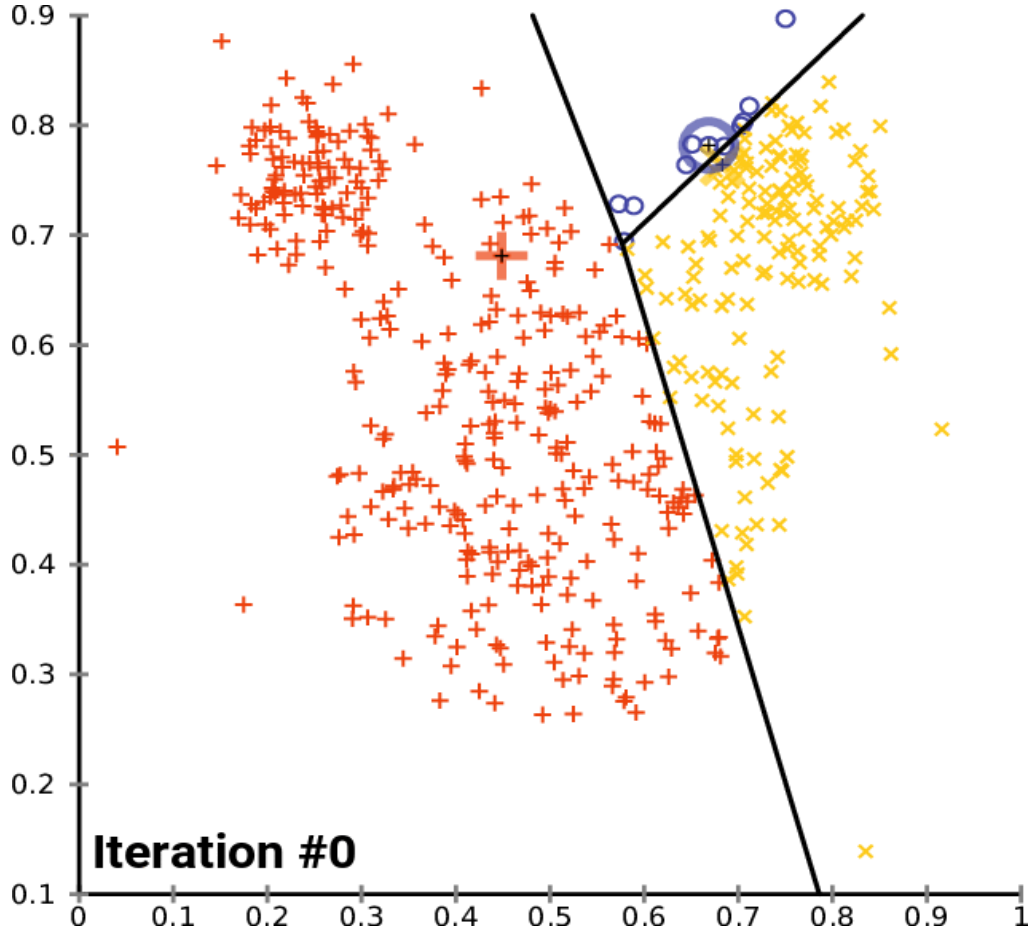
- Use **Principal Componant Analysis (PCA)** to visualize the 9 variables (9-dimensional) to a plot (2-dimensional)

Basic idea: Construct principal components from the features x_i.
The principal components (PC) are uncorrelated linear combinations, whose variances are as large as possible.



Clustering algorithms:

- K-means



Clustering algorithms:

Method name	Parameters	Scalability	Usecase	Geometry (metric used)
K-Means	number of clusters	Very large `n samples`, medium `n_clusters` with MiniBatch code	General-purpose, even cluster size, flat geometry, not too many clusters, inductive	Distances between points
Affinity propagation	damping, sample preference	Not scalable with n_samples	Many clusters, uneven cluster size, non-flat geometry, inductive	Graph distance (e.g. nearest-neighbor graph)
Mean-shift	bandwidth	Not scalable with `n_samples`	Many clusters, uneven cluster size, non-flat geometry, inductive	Distances between points
Agglomerative clustering	number of clusters or distance threshold, linkage type, distance	Large `n samples` and `n_clusters`	Many clusters, possibly connectivity constraints, non Euclidean distances, transductive	Any pairwise distance

Process of Data analytics

What are the steps of data analytics?

Clustering metrics:

How do you know how many clusters are present or how good your clustering results are?

- Determine the **within-cluster sum of squares (WCSS)**

$$\sum_{k=1}^{K} \sum_{x_i \in C_k} (x_i - \mu_k)^2 ,$$

where $\boldsymbol{\mu_k}$ is the center of the k-th cluster, $\boldsymbol{C_k}$, and $\boldsymbol{x_i}$ are all the points inside this cluster.

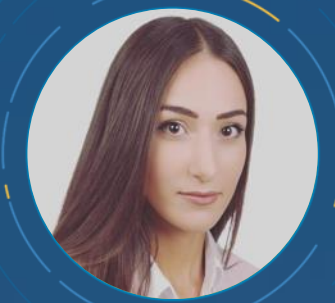
Basic idea: Measure of how tightly the data points are grouped within each clusters.
Alternatives: Silhouette score, Rand Index, Adjusted Rand Index, …

What are the currently available self-paced labs?

Here are some of the self-paced labs we have developed.

LAB	DATA	TOOLS
Onboarding: Local Computer	/	• Conda • Boto3
Onboarding: Data Science Studio	/	• Conda • Boto3 • Mysql.connector • elasticsearch
Python: Data Exploration	**EU Open Data:** Belgium COVID-19	• Pandas DataFrame • Matplotlib
Machine Learning: Regression	**EU Open Data:** Gender Equality Index Dataset	• Scikit-learn
Machine Learning: Classification	**Kaggle:** Breast Cancer Wisconsin Dataset	• Scikit-learn
Machine Learning: Clustering	**Kaggle:** Country Dataset	• Scikit-learn
Apache Spark: Data Exploration	**EU Open Data:** Belgium COVID-19	• PySpark DataFrame • Matplotlib
Apache Hive: Data Exploration	**EU Open Data:** Belgium COVID-19	• HiveQL • Matplotlib

CEF Digital
Connecting Europe



Francesca Vella
Pilot Request Manager
BDTI Team

Support & How to apply (10 min)

How BDTI will support you throughout your journey

Discover

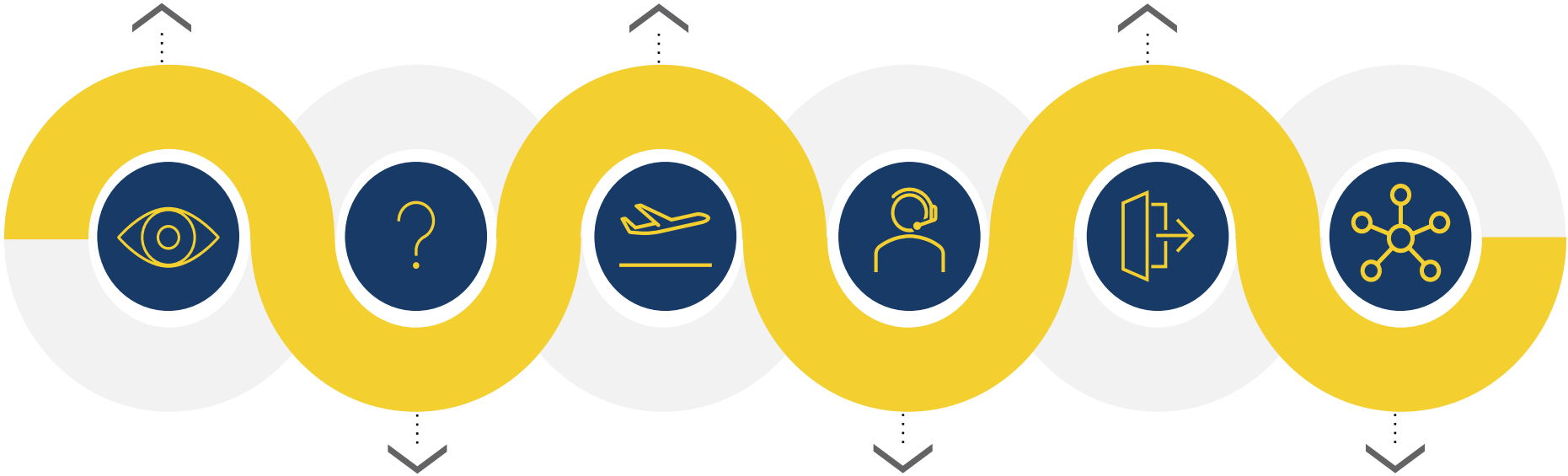
Help users to discover the services offered by BDTI and how they leverage insights from data to support the policy-making process, and support the onboarding process of stakeholders.

Start

Support the set up of the big data pilot, enabling users to benefit from the service desk, advisory, knowledge base and community, and support the integration of the stakeholder's data. A set of transparent tutorials, the self-paced-labs, are made available to help non-technical pilot members to start experimenting with their data in BDTI.

Finish

Support the in-house implementation of BDTI at the end of the pilot project.



Evaluate

Help users to identify the specific business needs and translate these into requirements to design the BDTI pilot taking into account the techniques, supporting software packages, infrastructure requirements and reporting tools needed, and choose the BDTI offering to use from among open source and commercial solutions. With this aim, BDTI provides users with catalogues of open data sources and analytics software that can be used in the pilots' analysis.

Use

Support the implementation of the big data pilot and provide technical assistance during the pilot.

Share

Support users in sharing their results in the BDTI Community Portal and in presenting the pilot highlights through a dedicated workshop.

How to get started with BDTI

Get familiar with BDTI services

Big Data Test Infrastructure offers a safe environment to experiment with Big Data technologies and to prototype Big Data solutions before deploying them in your own production environment

Define your data analytics use case

Build your BDTI use case. The support team is available to guide the users through the process of defining their organisation's use case, as well as to clarify any preliminary questions

Submit your "BDTI PILOT request"

Request to use BDTI by submitting an **online form.** The user provides information on the online form in order to clarify the pilot objective (general summary, short description and any supporting evidence)

Elaborate your business and technical need

The functional team will work with you on the elaboration of your business case to ensure that it fits within the CEF requirements, while our technical team will design your BDTI test environment which is tailored for your specific technical needs

Pilot project approval

The European Commission will give a final validation to run the pilot project.

Test environment set up

The BDTI technical team will set-up your tailored cloud-based data analytics environment so that you can start with your big data experiments.

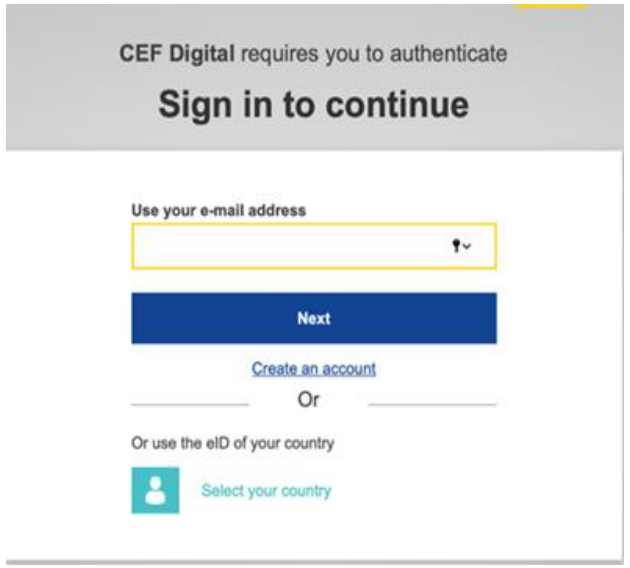
Start the pilot and gain insight

The users can experiment with data on the test infrastructure provided by BDTI and share any lessons learned on our Community portal

How to submit a BDTI pilot request (1/2)

1 Access to the **BDTI homepage** (https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Big+Data+Test+Infrastructure) and click on the "**REQUEST BDTI PILOT**" section.

2 To access the BDTI service desk, you will need to create an **EU login account**, entering your e-mail address and filling in the form with the required information.



How to submit a BDTI pilot request (2/2)

3 Once you have filled in the form, you will receive an email to set your EU **account password** and complete your registration.

4 Now you will be able to access to the BDTI service desk, inserting your password, putting **"Password"** as verification method and clicking on **Sign in**.

5 Once registration is complete, click on **BDTI PILOT REQUEST**, and fill in the form to submit your request.

CEF Digital Connecting Europe

Use cases acceptance criteria

Business criteria

- **Potential users**: Member State or public administration at **national** level, **regional** or **local** level
- **Clear value added**: Business and technical
- **Clear contact point** for the entire pilot

Functional criteria

- **Pilot duration**: 6 months
- **Pilot use cases**: (only use case in scope)
- **Resource usage limit**: based on CEF budget
- **Skills/Maturity level**: adequate skilled resources and/or level of maturity on the big data subject
- Pilot BDTI **geographical distribution/ resource allocation**

Case studies of ongoing pilots

Conselleria de Sanitat (CS)

Conselleria de Sanitat (CS) is the Health Public Administration, belonging to the Comunidad Valenciana (CV) Regional Government and it provides health services for all **5.2 million people** in the region. They needed a tool capable of analysing and synthetising the huge quantity of scientific clinical articles coming from different sources: PubMed.gov (more than 30M, and 1M coming every year) and the 100.000 + clinical articles Covid-19 related generated in the first 6 months of pandemic.

Municipality of Milan

The Municipality of Milan wants to analyse the **movements of people before and during the lock-down period**, in order to predict the future flow of the citizens and assess which areas will be more impacted by the full release of the current restrictions related to Covid-19.The available data collected from Telco operators were used to create a dashboard showing the most crowded areas of the city in the period before, during and after the lockdown.

.

City of Florence

The main goal of the Municipality is to perform a **cross correlation between the multiple datasets** available within the city to understand how people were and are moving between the different districts, to then derive precious insights about the most and the less crowded neighborhoods during and after the lockdown and about **how services can be relaunched to foster cultural activities and events.**

How BDTI is helping

BDTI is supporting Concelleria de Sanitat with advanced **data visualization** and **text mining** tools to help **extracting the knowledge contained in the documents**, supporting clinicians and managers in their clinical practices, management process and day-to-day work in fighting the virus.

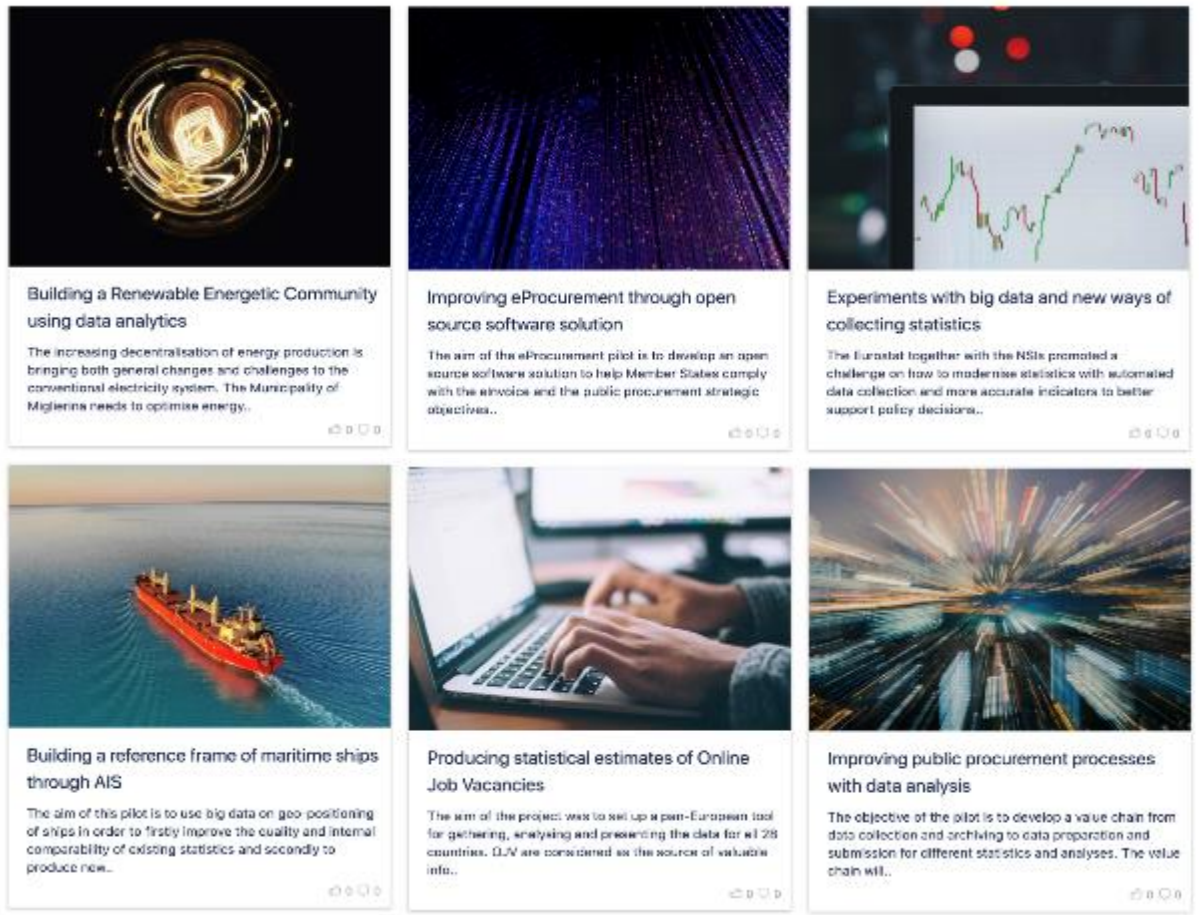
BDTI is supporting the Municipality of Milan with a scalable virtual environment capable to perform **descriptive, predictive and time-series analysis**. Inside the platform, Telco data can be correlated with on premises data about commercial activities, schools and universities to analyse the future flows of the citizens.

BDTI is supporting the City of Florence with predictive, descriptive and time-series analysis on multiple datasets collected **before, during and after the Covid-19 pandemic** such as: public wifi sensors, parking and geo-referenced data of people movements (i.e. tourists).

Want to learn more?

If you want to learn more about BDTI pilots, please visit **BDTI Community Portal**

https://ec.europa.eu/cefdigital/wiki/display/BDUC/Big+Data+User+Community+Home



Ready to get started?

Reach out to us to learn more!

Visit us at https://ec.europa.eu/cefdigital/bdti

Follow us on social media

@Connecting_EU

Connecting Europe

Connecting Europe

Drop us a line

CEF-BDTI-SUPPORT@ec.europa.eu

Q&As (10 min)

General BDTI service offering