Accenture Labs

Politecnico di Torino

APPLIED DATA SCIENCE PROJECT PROPOSAL

2021

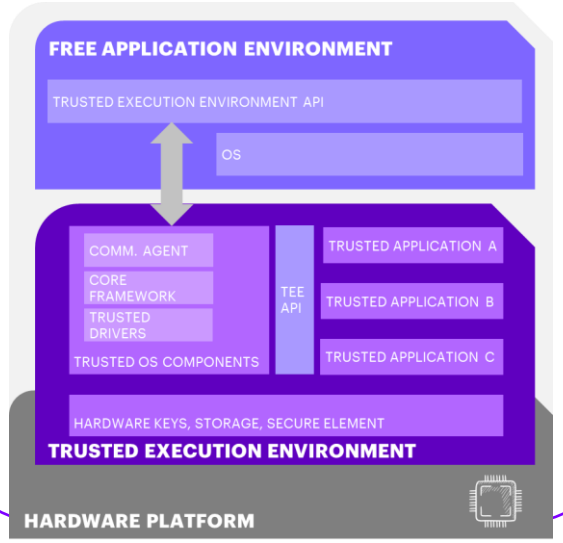
APPLIED DATA SCIENCE PROJECT PROPOSAL

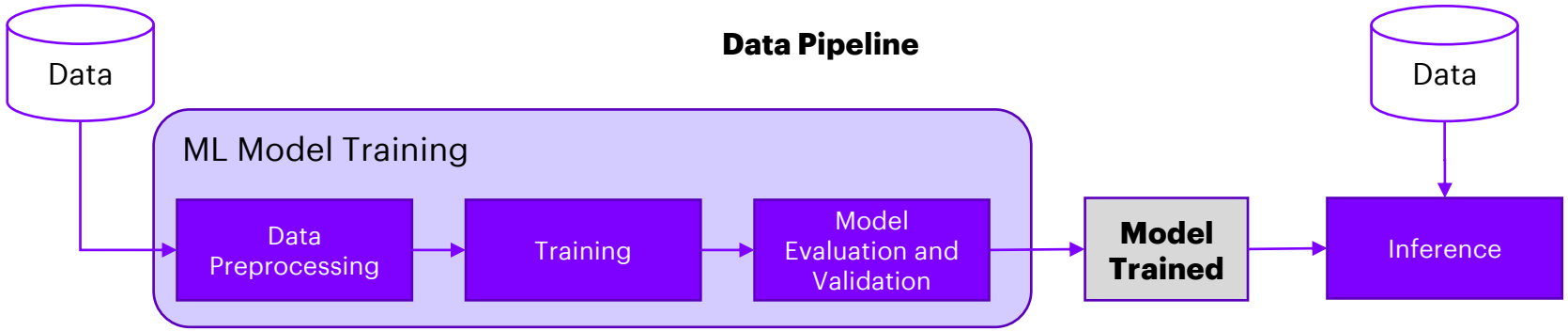
PROJECT: PRIVACY PRESERVING MACHINE LEARNING WITH SECURE SERVICE FOR INFERENCING 1/2

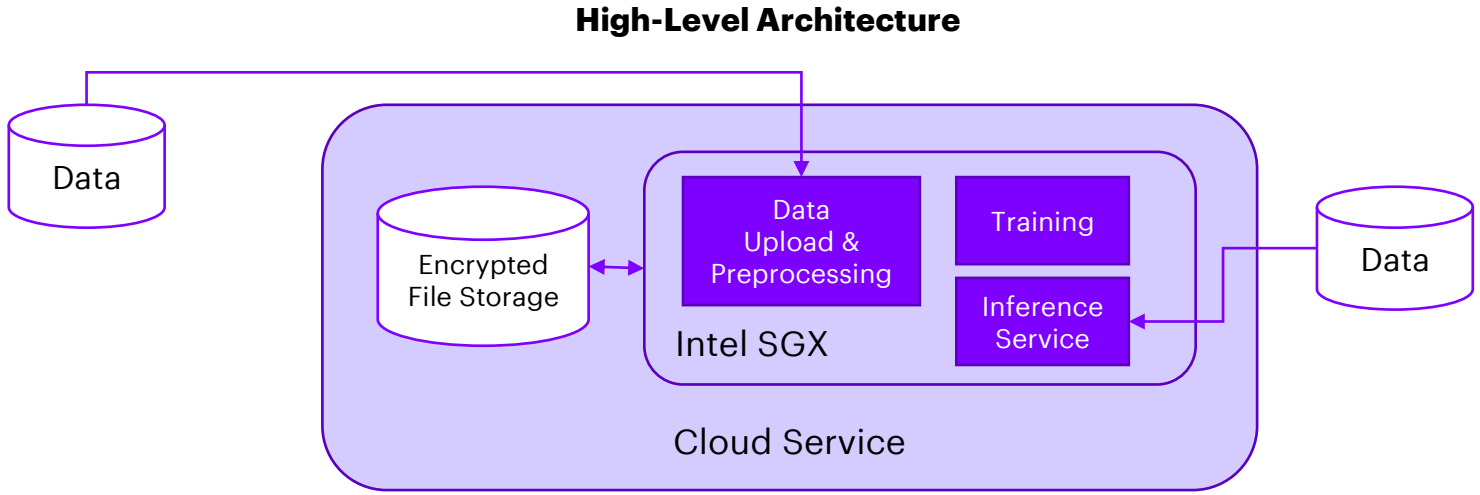
Scope: Build a Machine Learning Data Pipeline (from Data Preparation, to training, to inferencing) within a Trusted Execution Environment to enable data Privacy at all stages of the process.

What is a Trusted Execution Environment?

A **Trusted Execution Environment**, or **Secure Enclave** as they are sometimes known, is an environment with special **hardware modules** that allow for **data processing within hardware-provided, encrypted private memory areas directly on the microprocessor chip only accessible to the running process.**







APPLIED DATA SCIENCE PROJECT PROPOSAL

PROJECT: PRIVACY PRESERVING MACHINE LEARNING WITH SECURE SERVICE FOR INFERENCING 2/2

Scope: Build a Machine Learning Data Pipeline (from Data Preparation, to training, to inferencing) within a Trusted Execution Environment to enable data Privacy at all stage of the process.

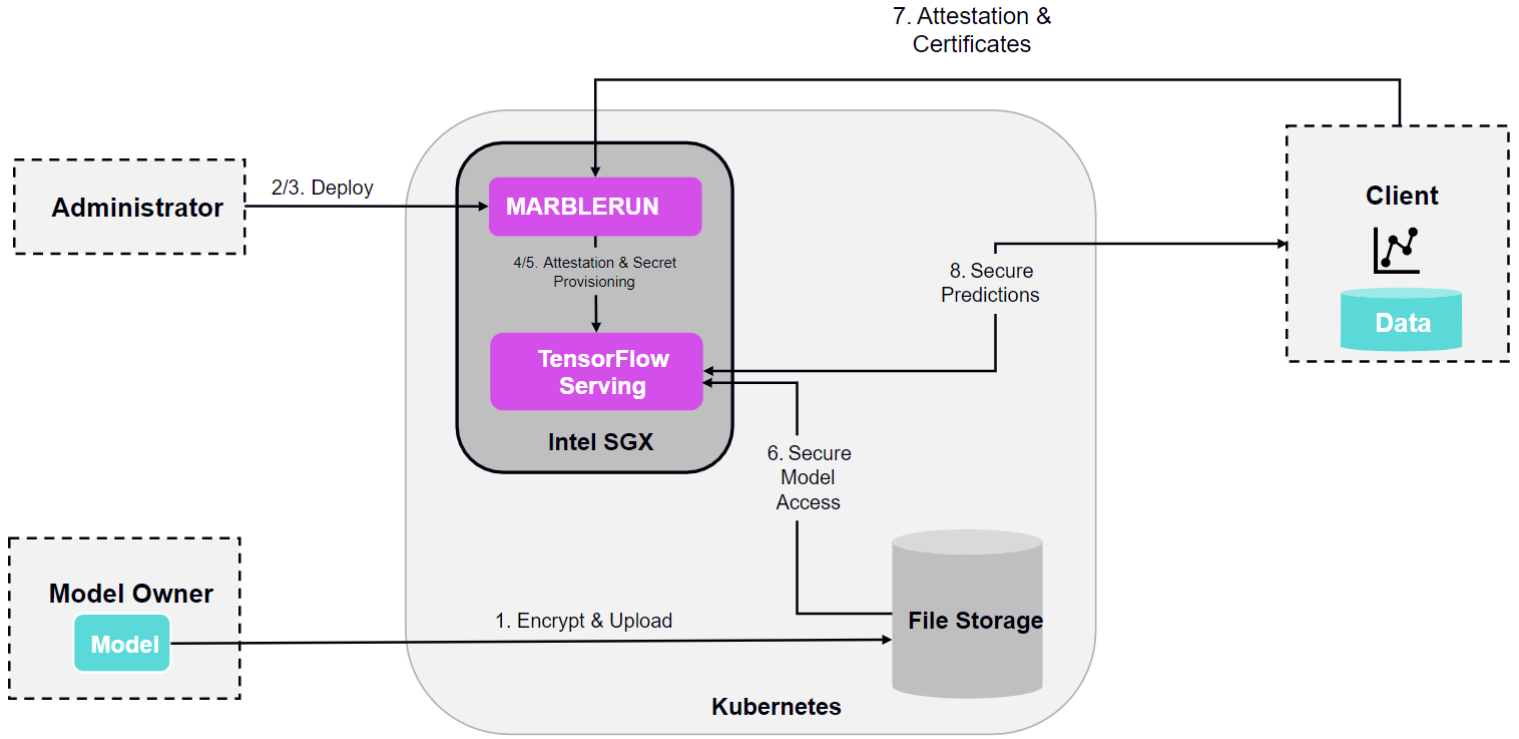
An example to start with

We propose to expand an existing example provided by the startup Edgeless Systems. The example implements an inference service that runs in a Secure Enclave. The example demonstrates how it is possible to protect the confidentiality of the Model and the Client Input's data while still allowing the inferencing and collaboration between the two parties.

Edgeless Systems is a German startup that provides open-source tools to develop TEE apps:

EGo: It enables running Go apps in Intel® SGX enclaves
MarbleRun: It facilitates the deployment, scale, and verification of SGX-based apps on Kubernetes.
Edgeless DB: It is a full SQL database that runs in a confidential computing environment.

EDGELESS SYSTEMS



Github: https://github.com/edgelesssys/marblerun-tensorflow-demo
Blog Post: https://blog.edgeless.systems/confidential-multi-stakeholder-machine-learning-2292f842e95a

APPLIED DATA SCIENCE PROJECT PROPOSAL

VALUE-DRIVEN PROJECT

Why is the project conducted?

The Sophia Antipolis **Systems & Platforms** team within the Accenture Labs is focusing on **Privacy Preserving Computation** technologies to enable **Data Collaboration in Multi-party Systems.**

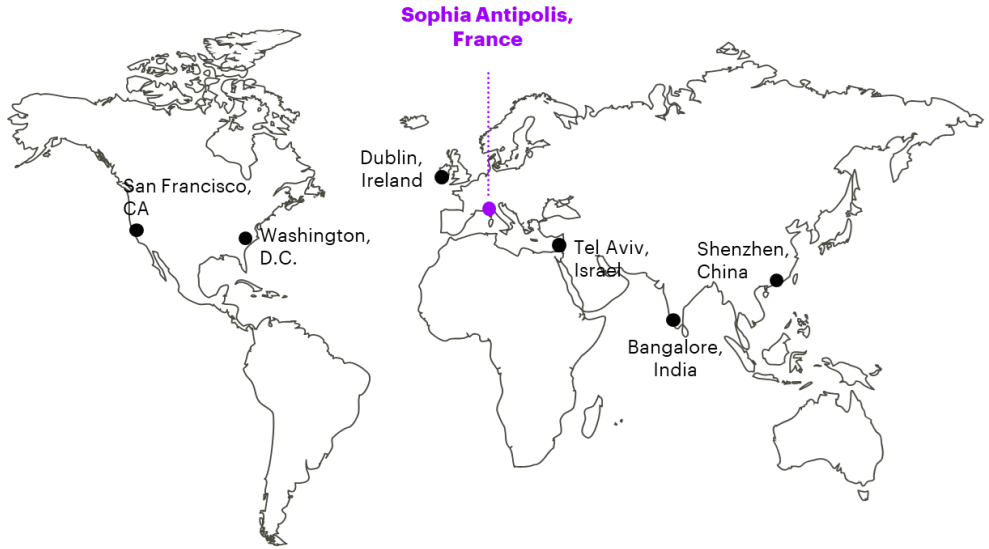
Trusted Execution Environments is one of the focus of the S&P team. This project aims to **expand knowledge on Edgeless Systems solution** and **build a new demonstrator** to promote these concept with Accenture Client's.

Who are the Stakeholders?

Accenture Labs **incubates and prototypes new concepts through applied R&D projects** that are expected to have a significant strategic impact on clients' businesses. Our dedicated team of technologists and researchers work with leaders across the company to invest in, incubate and deliver breakthrough ideas and solutions that help our clients create new sources of business advantage.

Accenture Labs are located in seven key research hubs around the world: Silicon Valley, CA; Sophia Antipolis, France; Arlington, Virginia; Beijing, China; Bangalore, India; Herzliya, Israel and Dublin, Ireland

Accenture Labs



APPLIED DATA SCIENCE PROJECT PROPOSAL

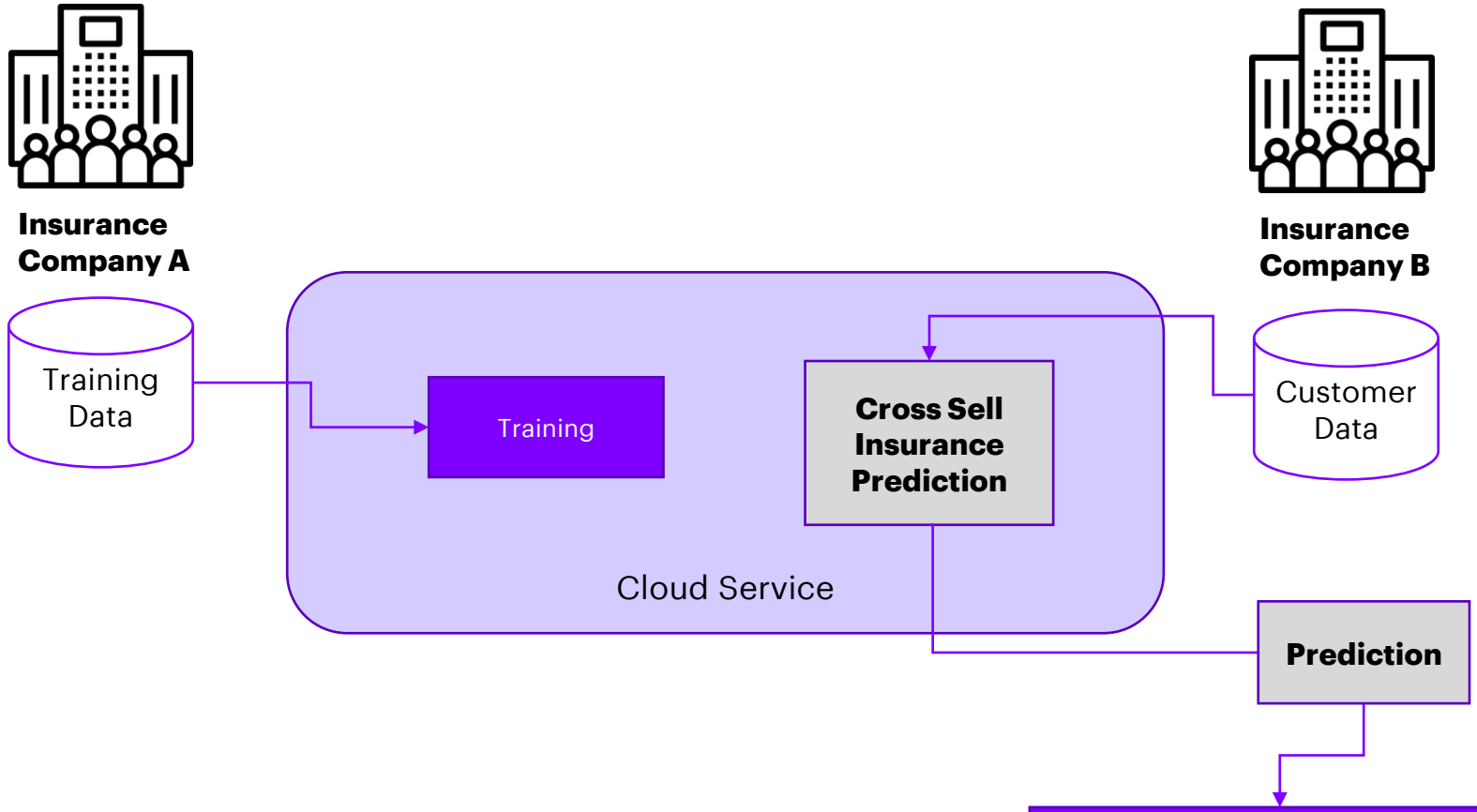
DATA (TBC): HEALTH INSURANCE CROSS SELL PREDICTION

Scenario

A Health Insurance company wants to build a **predictive model** to determine if their Health insurance policyholders would be also interested in a Vehicle Insurance.

The Health insurance company would also like to **monetize its model** with other insurance companies that could use the model to target new clients.

Secure Enclave allows companies to **outsource computation on the cloud** while protecting their privacy.



Kaggle Dataset:
https://www.kaggle.com/anmolkumar/health-insurance-cross-sell-prediction/tasks?taskId=2055

Insurance Marketing Teams can reach out to the right costumers and optimize their business model and revenue

APPLIED DATA SCIENCE PROJECT PROPOSAL

PRIVACY-PRESERVING COMPUTATION AND UN SUSTAINABLE DEVELOPMENT GOALS

The **data revolution** was recognized as an **enabler of the Sustainable Development Goals**. The 2030 Agenda asserts that "**Quality, accessible, timely and reliable disaggregates data will be needed to help with the measurement of progress (SGDs).**

There are legitimate concerns regarding **risks associated with handling and processing of big data**. These concerns continue to complicate efforts **to develop standardized and scalable approaches to risk management and data access**. A coordinated approach is required to ensure the emergence of frameworks **for safe and responsible use of big data f**or the achievement of the 2030 Agenda.

1 NO POVERTY
Spending patterns on mobile phone services can provide proxy indicators of income levels

2 ZERO HUNGER
Crowdsourcing or tracking of food prices listed online can help monitor food security in near real-time

3 GOOD HEALTH AND WELL-BEING
Mapping the movement of mobile phone users can help predict the spread of infectious diseases

4 QUALITY EDUCATION
Citizen reporting can reveal reasons for student drop-out rates

5 GENDER EQUALITY
Analysis of financial transactions can reveal the spending patterns and different impacts of economic shocks on men and women

6 CLEAN WATER AND SANITATION
Sensors connected to water pumps can track access to clean water

7 AFFORDABLE AND CLEAN ENERGY
Smart metering allows utility companies to increase or restrict the flow of electricity, gas or water to reduce waste and ensure adequate supply at peak periods

8 DECENT WORK AND ECONOMIC GROWTH
Patterns in global postal traffic can provide indicators such as economic growth, remittances, trade and GDP

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
Data from GPS devices can be used for traffic control and to improve public transport

10 REDUCED INEQUALITY
Speech-to-text analytics on local radio content can reveal discrimination concerns and support policy response

11 SUSTAINABLE CITIES AND COMMUNITIES
Satellite remote sensing can track encroachment on public land or spaces such as parks and forests

12 RESPONSIBLE CONSUMPTION AND PRODUCTION
Online search patterns or e-commerce transactions can reveal the pace of transition to energy efficient products

13 CLIMATE ACTION
Combining satellite imagery, crowd-sourced witness accounts and open data can help track deforestation

14 LIFE BELOW WATER
Maritime vessel tracking data can reveal illegal, unregulated and unreported fishing activities

15 LIFE ON LAND
Social media monitoring can support disaster management with real-time information on victim location, effects and strength of forest fires or haze

16 PEACE, JUSTICE AND STRONG INSTITUTIONS
Sentiment analysis of social media can reveal public opinion on effective governance, public service delivery or human rights

17 PARTNERSHIPS FOR THE GOALS
Partnerships to enable the combining of statistics, mobile and internet data can provide a better and real-time understanding of today's hyper-connected world

Data Privacy, Ethics and Protection (UN 2030 Agenda)

TASK

1. **Dataset Analysis**
2. **ML Model Evaluation and Validation**
3. **Test Edgeless Systems Example**
4. **Solution Architecture with Privacy Preserving Computation/Edgless DB**
5. **Implementation of ML Training with Edgeless System (Leveraging Edgeless DB, Ego and MarbleRun)**
6. **Implementation of Inference Service (Client to be protected, Service to run in the Enclave)**
7. **Cloud Deployment**
8. **UI for Inferencing service (to run in the Secure Enclave)**

N.B: The list of task can be reviewed according progress during the execution of the project.

APPLIED DATA SCIENCE PROJECT PROPOSAL

LIGHT MENTORING

- **Biweekly Call during the duration of the course**
- **Off-line support via mail**

- **(To Be Confirmed) Introduction to Edgeless Systems by the Edgeless Systems team**

Anh-Dung Le
anh-dung.le@accenture.com

Luca Schiatti
luca.schiatti@accenture.com

Giuseppe Giordano
giuseppe.giordano@accenture.com

Accenture Labs

EDGELESS SYSTEMS