Pooling data across banks to better fight financial crime

By combining their transactions data, financial institutions could deepen their knowledge to improve the detection of suspicious activities, terrorism financing, fraudulent and money laundering operations. But sharing critical data among banks raises significant compliance and security risks.

This challenge is one of the main areas of investment identified by the Financial Action Task Force (FATF) to improve anti-money laundering and terrorist financing. The French banking supervisor (ACPR) selected Sarus with Microsoft and EY, to address this high stake goal.

100% of data utility preserved

50+ features in AML model

0 transactions ever exposed

0 code change compared to having direct data access

"As AML data-scientists, we were able to use these technologies successfully, easily and quickly, without any impact neither on our way of working nor on integration within our processes and systems."

– **Chadi Sassine**, Partner FInancial Services, EY

SOLUTION: CONFIDENTIAL COMPUTING WITH PRIVATE ACCESS LAYER

The consortium proposed a solution that combined the state-of-the-art Microsoft Azure confidential computing environment with a Sarus private access layer. EY analysts and data scientists could leverage the sensitive data from the Sarus API without ever being able to extract confidential information.

This architecture addressed both input and output privacy risks. It enabled EY data science team to bring in their strong expertise in AML modeling and develop the most advanced detection models, whether rule-based or learning-based models.

SARUS: PRIVACY-BY-DESIGN DATA ACCESS

With Sarus, datasets become instantly available in a compliant and secure way. Data practitioners carry out analyses on sensitive data without ever seeing it. Data security is solved at scale by implementing the highest data protection standards: *Differential Privacy*.

NOVEL ARCHITECTURE, NO IMPACT ON DATA SCIENCE EXPERIENCE

The consortium introduced a novel architecture that combined the confidential computing solutions and the privacy-preserving proxy. Confidential computing made all data flows encrypted so that neither the cloud provider, nor the Sarus team or the participating banks could ever access transaction data. Data was encrypted at rest and in transit and eventually stored in a Azure SQL DB running on a confidential VM. By using architecture-as-code principles, the Sarus instance was the only module with the ability to query the confidential enclave, which was enforced by attestation. The Sarus proxy ensured that queries from the analysts or developers were unable to extract transaction-level information. This guarantee stems from the strict application of differential privacy.

Though data processing jobs were handled by Sarus and ran in a confidential VM, the analyst experience was fully preserved. EY data scientists implemented AML models on the pooled data using the exact same libraries and internal practices they would have used should they access the data. To facilitate exploration and debugging, Sarus provided synthetic data versions of all remote tables. The generation of synthetic data was fully automated and did not require anyone accessing the data. The generation process also relied on differential privacy to protect against all inference attack based on the samples.

RESULTS: BETTER DETECTION RATES

The consortium combined the state-of-the-art of confidential computing and privacy techniques to successfully develop a powerful AML detection model. At no point, transaction data was available in clear to any party, showing that financial institutions can pool their data in a safe and compliant manner.

This achievement can be delivered at the scale of Azure cloud and while preserving the experience of data scientists, which is crucial to push the best models to production. The EY data science team could design, test, and deploy a rules-based and a machine-learning based model that considerably improved detection rates while reducing false positives.