



# **Anti-Money Laundering with Deep Learning on Hopsworks**



# AML: Anti-Money Laundering

**Money laundering has been a global news scandal during 2018 and 2019. Despite the best efforts by regulators and banks, existing techniques to identify and stop money laundering, so-called anti-money laundering (AML), have been shown to be inadequate.**

Banks have traditionally moved slowly with technology, for good reasons, but money launderers do not. Existing systems for identifying money laundering are rules-based systems, designed and managed by domain experts, but in this white paper we outline a new approach based on artificial intelligence (AI), and, in particular, a technique called *deep learning* that has been shown to have superhuman capabilities in identifying patterns in data.

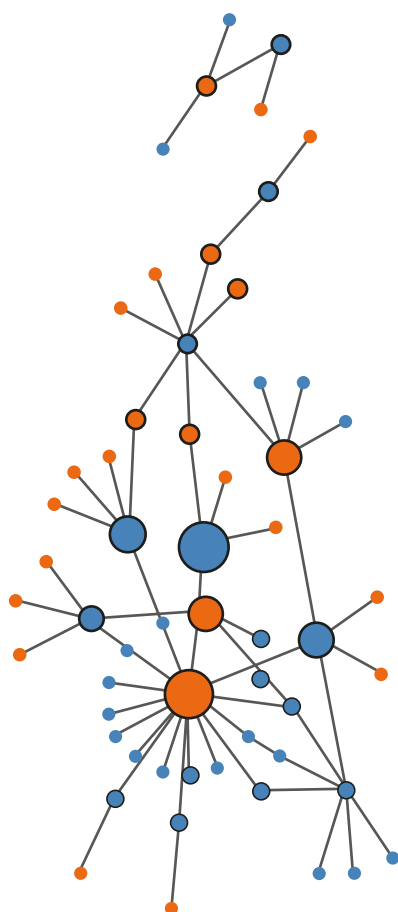
The current rules-based AML technology is costly in terms of time and resources. Existing rule-based systems for identifying transactions that may involve money laundering consist of thousands of rules. They generate an alert when a rule matches for a transaction - suspecting the transaction to involve money laundering. These systems generate a huge numbers of *false-positive alerts* (alerts where the transaction did not involve money laundering) that take time and money to run down. Rule-based approaches are not capable of detecting changing threats, as the rules are not able to generalize to capture similar but slightly modified threats. As such, money launderers constantly try to find new ways around the rules, and after the fact, banks add new rules to catch newly detected threats. Recent AML scandals have caused regulators to increase pressure on banks to increase their spending on AML and compliance – which can only mean more alerts and more costs in handling more false-positives, while the actual threats are modifying their behavior to get around the rules designed to catch them.

# Promise of Deep Learning for AML

The key insight of deep learning for AML is that deep neural networks (DNNs) can generalize from training data to identify patterns in transactions that are indicative of fraud. That is, having been shown some patterns in real money laundering situations, DNNs can identify similar and modified patterns as also AML. This makes it harder for money launderers to make small changes in how they launder the money while coming under the radar of AML alerts (not triggering rules).

Deep learning can also model transaction data from different perspectives, giving different roads for identifying money laundering. One perspective is time: money laundering can sometimes use semi-dormant or proxy accounts to transfer batches of money, but deep learning can identify unusual patterns of transactions over different periods of time, such as over a given week or month or quarter.

Another useful perspective on transaction data is to represent transactions between customers in a graph, where the customers are nodes in the graph (circles) and transactions between customers are the edges (connections) in the graph – see the figure on the left. The graph perspective enables us to identify a recurrent pattern in money laundering - the so-called dandelion pattern, where suspicious customers often have lots of connections going to either customers outside of the bank or connections to many customers who, in turn, have follow-on connections. Slight variations of these patterns can trick rule-based systems, but are detected by deep learning systems. In the graph perspective, we have successfully used a new technique known as *graph embeddings* to capture a snapshot of parts of the graph, so that deep learning can learn the suspicious patterns.



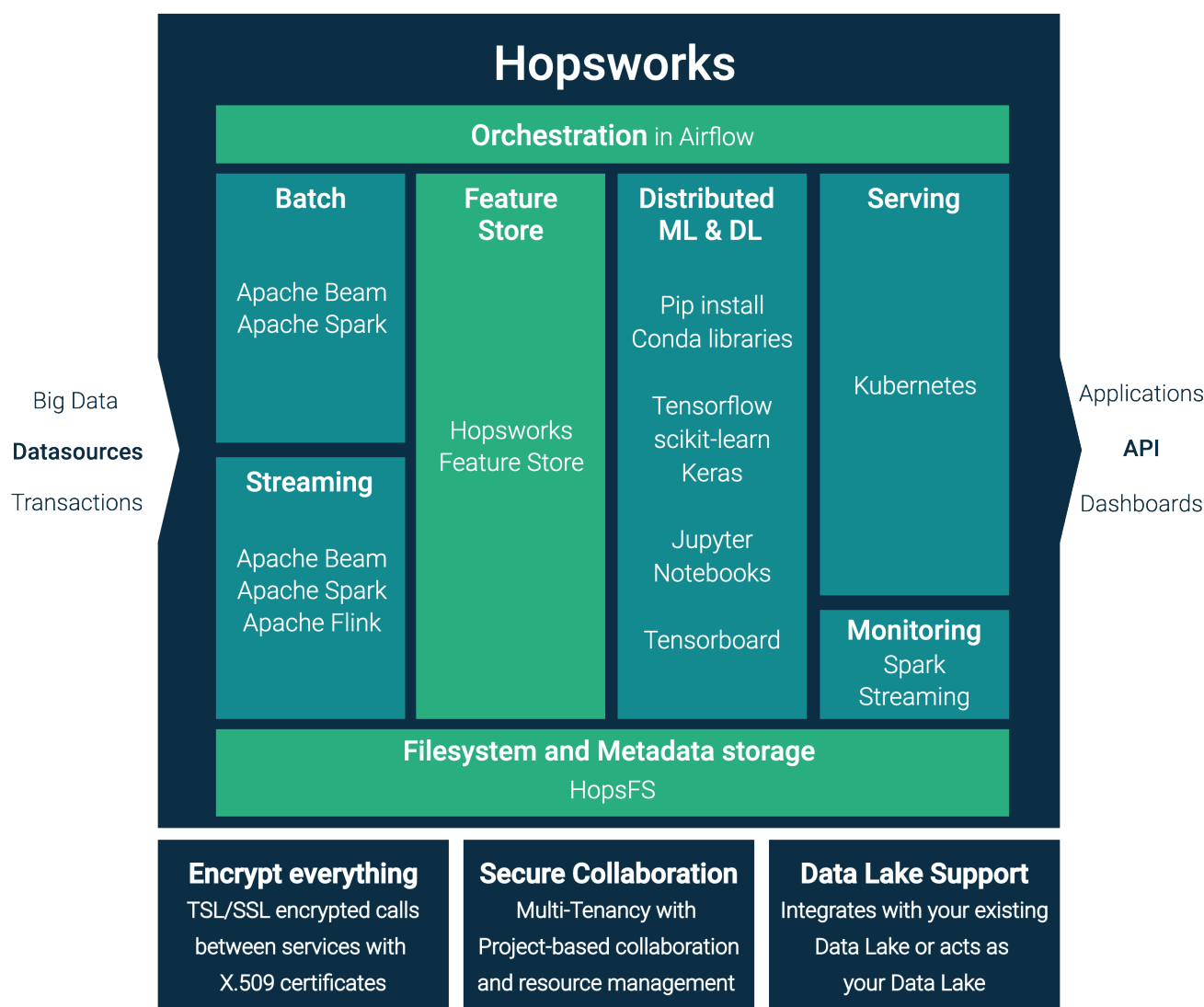


Figure 1: End-to-End Machine Learning Pipelines with Hopsworks

# Deep Learning on Hopsworks for AML

Logical Clocks have experience with Nordic banks in enhancing existing rule-based systems through deep learning (supervised learning) using the Hopsworks platform. The Hopsworks platform is used to both develop and operate machine learning and deep learning models, at scale. For AML, the scale problem is particularly acute, as banks typically have billions of historical transactions over many years that could be used to train DNNs using supervised learning on graphical processing units (GPUs).

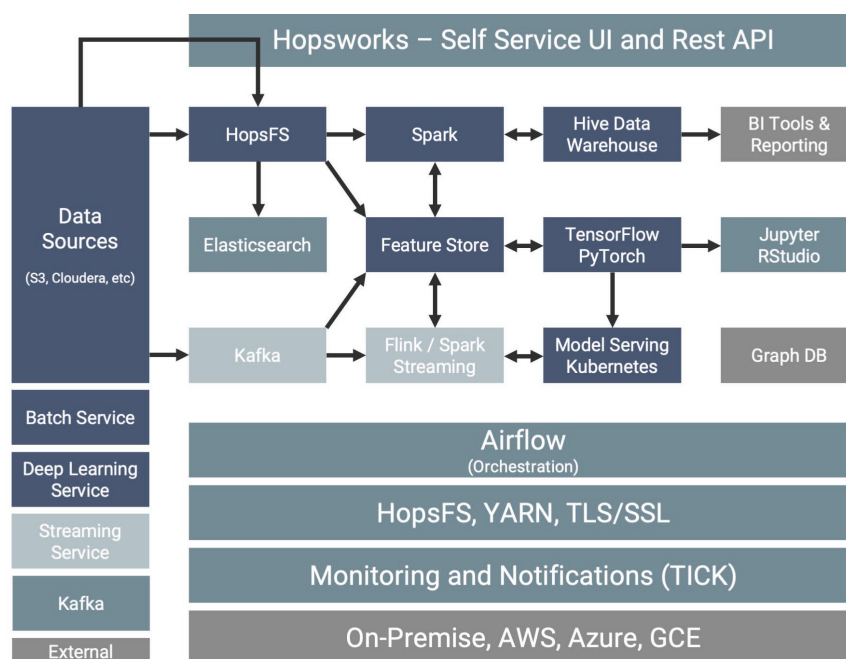


Figure 2: Hopsworks Data Platform for AI

Hopsworks integrates seamlessly with existing data lakes (Cloudera/Hortonworks/S3/Azure) and can easily scale to train models on massive amounts of data using as many GPUs as are available for training. This makes the problem tractable of training large models on large amounts of data for AML. It also makes data scientists more productive through reducing the time needed to find good hyperparameters – Hopsworks supports distributed hyperparameter tuning spread across many GPUs. Hopsworks can also reduce training time from weeks or months for large AML models to hours or days using distributed training over many GPUs – with CollectiveAllReduce on TensorFlow or PyTorch.

Another challenge that Logical Clocks has addressed successfully on the Hopsworks platform is handling the massive class imbalance in the raw transaction data - there are a very small relative number of transactions are reported to authorities as money laundering compared to the total number of transactions. We have experience developing methods for data synthesis that have worked effectively for graph embeddings and temporal features. We also have experience at automatically cleaning/preparing/enriching data, and making features reusable by other groups using our Feature Store. The Feature Store acts as an effective API between team members who are working on data engineering versus those working on Data Science (model building, training, and evaluation).



the creators of

# HOPSWORKS

**Reach out to us!**

[www.logicalclocks.com](http://www.logicalclocks.com)

[info@logicalclocks.com](mailto:info@logicalclocks.com)

**.. or visit us at one of our offices:**

**Silicon Valley Office**

470 Ramona St  
Palo Alto,  
94301 California  
USA

**Stockholm Office**

Box 1263,  
Isafjordsgatan 22  
164 40, Kista  
Sweden

**UK Office**

IDEALondon,  
69 Wilson St  
EC2A2BB, London  
UK

Anti-Money Laundering with  
Deep Learning on Hopsworks