

Spark'ing an Anti-Money Laundering Revolution

Katie Levans | Koert Kuipers | Tresata



Overview

The Thin Line Between Good & Evil

Why Money Laundering is Hard to Stop

Lots of Rules, Not Enough Tools

Why AML is Top of Mind (But Hard to Do Well)

Spark'ing a Revolution

Tresata + Databricks Solution



TOO BIG

\$2 TRILLION PROBLEM

A group of women and children, likely in a conflict-affected region, are shown in a crowd. They are wearing headscarves in various colors like pink, grey, purple, and blue. Many of them have their hands raised or are clapping, suggesting a moment of celebration or protest. The background is slightly blurred, focusing attention on the people in the foreground.

TOO BROAD

FOOD OR GUNS



TOO DUMB

98% false positives



TIMES HAVE CHANGED

AML IS A NETWORK PROBLEM

Money Laundering Prevention

Customer Due Diligence (CDD)

Requirements to Know Your Customer by naming each account holder & proving identity

Reporting

Requirements for reporting of information to regulators

Regulation & Supervision

Countless rules and governing bodies to keep up with

Sanctions

Financial punishment for failing to uphold the first three measures

AML Fines

\$10 million Ocean Bank (2011)

\$1.2 billion HSBC (2012)

\$2 billion JP Morgan Chase (2014)

Problems with Current AML Tech

Poor entity resolution and identification

Tedious manual investigation

Lack of automation

Inaccuracy

SPARK'ING A REVOLUTION

 tresata +  databricks™



REALLY FAST

A close-up photograph of a dartboard. Three darts with yellow and black striped flights are clustered tightly in the center bullseye. The dartboard features a standard design with alternating black and white segments, and concentric rings of red and green. The text "REALLY PRECISE" is overlaid in large, white, sans-serif capital letters across the center of the board. On the right side, a small logo for the "BRITISH DARTS ORGANIZATION" is visible, along with some handwritten numbers like "12", "10", and "3".

REALLY PRECISE



REALLY SMART

Tresata's AML Engine

Our needs for our core graph engine:

- Interactive

Most queries within few seconds | Large queries no more than 20 seconds

- Graph traversal

Property graph | Multi-hop | Shortest paths

- Search Engine

Not just point queries | Ability to become a graph search engine

- No Respect for Size

>50mm entities | >500mm relations

Tresata's AML Engine

Using Spark to make it happen:

- **Interactive**

In memory RDDs | Optimized data structures

- **Graph traversal**

Co-partitioning of RDD's | Iteration comes easy | In-house DSL on top of RDD's

- **Search Engine**

Dataset Scans | RDD batch processing (in milliseconds)

- **No Respect for Size**

Can fit 50MM nodes/ 500MM edges in memory on 10 nodes | 10X bigger if serialized in memory (3X slower though)

QUESTIONS?

www.tresata.com | katie@tresata.com | koert@tresata.com